Evaluation of Whizz Education: Maths-Whizz Intelligent Tutoring Programme
Evaluation Protocol

Evaluator (institution): NFER
Principal investigator: Stephen Welbourne

Evaluation summary

Project title	**Evaluation of Whizz Education: Maths-Whizz Intelligent Tutoring Programme**
Developer *(Institution)*	Whizz Education
Evaluator *(Institution)*	National Foundation for Educational Research
Principal investigator(s)	Stephen Welbourne
Protocol author(s)	Aarti Sahasranaman, Katherine Aston, Gemma Schwendel, Stephen Welbourne
Trial design	Two-arm interleaved randomised controlled trial with random allocation at school level
Trial type	Efficacy
Pupil age range and Key stage	5 – 9 years old at baseline; 6 – 10 years old at endpoint; KS1, KS2
Number of schools *(at design stage)*	64
Number of pupils *(at design stage)*	Approximately 9,570
Primary outcome measure and source	Maths attainment (Renaissance Star Maths assessment)
Secondary outcome measure and source	Mathematical self-perceptions, enjoyment of mathematics (Maths and Me survey)

Protocol version history

Version	Date	Reason for revision
1.2 [*latest*]		
1.1		
1.0 [*original*]		N/A

- Any changes to the design need to be discussed with the EEF Evaluation Manager and the developer team prior to any change(s) being finalised. Describe in the table above any agreed changes made to the trial design.

- Please ensure that any changes to the design of the trial that affect the analysis to be undertaken are also reflected in the SAP.

Contents

Tables

Figures

Study rationale and background

According to the Department for Education's (DfE) latest publication on attainment for the Key Stage 2 National Curriculum assessments in England, in the 2022-23 academic year, 53% of pupils from disadvantaged backgrounds met the expected standard in maths compared to 73% of pupils from non-disadvantaged backgrounds. The COVID-19 pandemic, while damaging to the maths learning of all pupils (Renaissance Learning and Education Policy Institute, 2021; EEF, 2022), has only further widened the attainment gap between disadvantaged pupils and their more affluent peers. A study commissioned by the Education Endowment Foundation (EEF) estimated that between the autumn term of 2019 and the end of the 2020-21 academic year, the maths attainment gap between disadvantaged pupils and their peers had widened by an equivalent of one additional month's progress (EEF, 2022b). Maths attainment at primary school is crucial because the links between Key Stage 2 maths performance and later educational achievement are particularly strong (Menzies, Ramaiah and Boulton, 2021). This means that pupils who perform poorly in maths at primary school are unlikely to be able to turn this around and perform better at the end of secondary school, with unfavourable downstream social and economic consequences (YouGov, 2014; Hodge, Little and Weldon, 2021).

Tutoring, when targeted to disadvantaged pupils, could provide an effective means to increase attainment and close the attainment gap between disadvantaged pupils and their peers. In fact, tutoring is being widely adopted as an academic catch-up intervention in response to the COVID-19 pandemic. The UK's DfE, for example, launched the National Tutoring Programme, with the stated aim of providing high quality tutoring to disadvantaged pupils in England (DfE, 2022). Evidence shows that one-to-one tutoring is effective at improving pupil outcomes with pupils making, on average, an additional five months' progress (Harrison and Higgins, 2023). Delivering one-to-one tuition is more expensive, and this is especially true when tutoring is delivered by qualified teachers. Small group tuition, in comparison, has a slightly smaller impact, with an average of an additional four months' progress (Harrison and Higgins, 2023) and may be a cost-effective solution for schools. However, the increasing demand for tutoring has not been commensurate with the supply of qualified tutors, particularly in maths.

Artificial intelligence (AI) enabled tutoring systems show promise both in terms of offering a more affordable tutoring option and sidestepping the requirement for a qualified human tutor. They also have the potential to support teachers who face the challenging and time-consuming task of teaching pupils with widely varying abilities within the same classroom. For instance, research by Whizz Education (developers of the Maths-Whizz programme) based on a representative sample of 400 UK primary classrooms using Maths-Whizz has shown that by upper primary there is a four-year knowledge gap between learners in a classroom (Whizz Education, 2021). AI-enabled adaptive tutoring systems can simulate human tutors by providing personalised tutoring, adapting to the individual needs of the pupil and providing instant feedback. A systematic review of maths apps for children aged four to seven found that personalised, adaptive features are more important to maximising children's maths attainment than non-adaptive approaches (Outhwaite et al., 2022), highlighting the promise of intelligent tutoring systems. However, evidence of the effectiveness of intelligent tutoring systems, particularly those supporting maths teaching, remains scant.

Maths-Whizz, developed and delivered by Whizz Education[1], is an EdTech programme consisting of an online intelligent tutoring system, a library of digital resources for teachers and live reports that support decision-making by teachers. Schools receive implementation support from Whizz Education, which includes training on the use of the platform, regular monitoring of platform usage and termly and annual progress reviews. Maths-Whizz is currently active in 21 countries and has reached over 9,000 schools and tutored 1.5 million pupils.

A quasi-experimental study of the impact of Maths-Whizz on maths attainment of pupils showed an effect size equivalent to an additional three months' progress (Mavrikis, Schlepps and Mubeen, 2018). The study involved 3,400 pupils in Mexico aged eight to nine years (fourth grade) who had access to the Maths-Whizz platform for seven weeks. This efficacy trial will be the first to evaluate the impact of the Maths-Whizz Intelligent Tutoring programme in English schools using a robust experimental design. This Randomised Controlled Trial (RCT) is larger in scope than the previous study and will involve pupils in Years 2, 3, 4 and 5 (ages 6 – 10) in 64 primary schools in England who will receive access to the Maths-Whizz programme for a whole academic year. Schools will be randomly assigned to either receive the intervention in Years 2 and 4 OR in Years 3 and 5. The year groups not receiving the intervention in each arm will serve as the control group for the year groups receiving the intervention in the other arm. Pupils will receive access to the Maths-Whizz programme during the 2024-25 academic year. The primary outcome is maths attainment measured using the Renaissance Star Maths assessment. The secondary outcome is pupil motivation towards maths learning measured using the Maths and Me survey. The integrated Implementation and Process Evaluation (IPE) will seek to explore how and in what circumstances Maths-Whizz impacts pupils and teachers. We will explore these questions using a combination of qualitative and quantitative methods. The qualitative methods will consist of interviews with teachers and senior leaders and focus groups with pupils while the quantitative methods will involve analysis of surveys of teachers and analysis of Maths-Whizz platform data. The cost evaluation will provide a robust estimate of the cost to schools per pupil per year of delivering the Maths-Whizz programme.

Intervention

Maths-Whizz is an adaptive virtual tutor developed by Whizz Education. This artificial intelligence (AI)-enabled tutor is designed for children between the ages of five and thirteen years. A detailed description of the intervention in the context of the TIDieR checklist is presented in the table below.

Table 1: TIDieR checklist for Maths-Whizz

TIDieR Item	Description
Brief Name	Maths-Whizz Intelligent Tutoring Programme
Why	One-to-one tutoring has been shown to be effective in improving pupil outcomes. AI-enabled tutoring platforms such as Maths-Whizz simulate human tutors by providing personalised tutoring, adapting to the individual needs of the pupil and providing instant feedback. Each pupil's journey on Maths-Whizz begins with an interactive and fully adaptive initial assessment to diagnose the gaps in their knowledge and determine their current ability

[1] https://www.whizzeducation.com/

	across several key topics in the maths curriculum. The tutor then creates a learning and support plan for each pupil tailored to their specific needs based on their 'Maths Age'. 'Maths Age' measures each pupil's maths ability against the expected level of an average pupil of their age. In addition to the overall 'Maths Age', the tutor will also calculate pupils' 'Maths Age' for each topic (called Topic Age) in the maths curriculum. Maths-Whizz provides personalised tutoring to pupils in Reception to Year 8 and covers topics ranging from place value and properties of numbers to shape and space and probability. The comprehensive list of 22 topics covered by Maths-Whizz can be found here. The topics for Reception to Year 6 are fully aligned to the National Curriculum learning objectives. Maths-Whizz is aligned to 96% of the learning objectives up to Year 8. The Maths-Whizz Tutor is underpinned by the Maths-Whizz curriculum which comprises 1200+ individual learning objectives covering the 5-13 age range. Learning objectives are derived from previous and current curricula frameworks, as well as Whizz's general understanding of mathematical concepts. To enable a curriculum structure, Whizz groups each lesson into one of 22 topics. Within a topic, lessons are strictly ordered. The exact position of a given lesson in its topic depends on the age at which students are expected to encounter that learning objective.
What (Materials)	All schools must purchase a School Solution which provides access to the Teachers' Resource, Teacher Platform for an unlimited number of teachers within the school and ongoing implementation and support. In addition to the School Solution, each school must also purchase individual subscriptions for each student who will access the Maths-Whizz platform. Parent access is provided via the student access so one parent will gain access to the parent platform for each individual student subscription activated. All schools participating in the trial will pay a subsidised fee of £500 to access both the School Solution and the Maths-Whizz programme for all participating pupils in two year groups (i.e., either Years 2 and 4 OR Years 3 and 5). All pupils will be provided with a login and password to access their personal Maths-Whizz accounts and the content within. This includes lessons for each learning objective, practice exercises, tests and Topic Challenges to review and consolidate their learning. Each Whizz learning objective is covered by a Whizz lesson. A Whizz lesson consists of three parts: • A Teaching Page, which introduces the underlying concept to students.

	• An interactive exercise (usually made up of ten questions along with help for when students get stuck). • A test, which mimics a worksheet and usually consists of 5 questions. Teachers will receive access to their own accounts where they will be able to view and monitor the progress of all pupils in their class. Teachers will also receive access to a digital library, Maths-Whizz Teachers' Resource containing over 1,200 instructional and assessment resources to supplement the Maths-Whizz tutor. These resources are designed to be used with the whole class or with individual students and are aligned to the National Curriculum. The types of resources provided include: • **Lessons** – engaging, animated and interactive exercises and tests, to use in class, organised by Topis or Year group to support lesson planning • **Activities** – a wide range of engaging problems, challenges, activities and games to complement learning in the maths lesson • **Worksheets** – includes printable worksheets to consolidate learning and support formative and summative assessments • **Enrichment sheets** – Rich tasks to develop mathematical reasoning, and problem-solving skills to develop fluency, application and to challenge pupils. • **Classroom resources** – a range of key resource and learning tools to support and enhance the learning of maths concepts In addition, using the comprehensive live learning data (automatically generated as pupils engage with Maths-Whizz) and the associated reporting platform, teachers can identify learning gaps and provide further targeted intervention where needed. Intuitive learning data can also inform decision-making for teaching and learning improvements, curriculum delivery and education strategy. Parents will receive access to a parent account where they can engage with their child's learning and monitor their child's progress and needs.
What (Procedures)	Following recruitment, Whizz Education will develop a bespoke implementation plan aligned to each school's goals and gain leadership support to ensure implementation is engrained within the school's core maths provision. To initiate this implementation planning process, Whizz Education will set up a 45 – 60 minute session with the school leadership and maths lead to determine the school's vision for success, long-term and short-term goals, agree training dates and action milestones and reinforce high-level expectations.

	All schools will receive training and onboarding from a Whizz Education Success Partner (ESP) to enable school leadership and teaching staff to facilitate pupil engagement with the tutor and empower staff to understand and act upon the associated learning analytics. • Session 1: This session includes expectation setting, basic knowledge of the Maths-Whizz platform (e.g. completing assessments, adaptive nature of the tool, what a typical lesson looks like and an overview of the teacher admin platform). It will be conducted as soon as possible upon starting in August/September 2024 with teachers of all participating classes and at least one member of the leadership team. • Session 2: This session is tailored around each school's needs. There is a focus on the teacher platform and teachers' data feed, using the Maths-Whizz platform to improve quality first teaching, a deeper dive into class reports, interpreting and actioning data, and how to use the tools such as Teachers' Resources and Topic Focus to support teaching. This 45-minute session with the maths lead (and teachers of participating classes subject to availability) will be conducted 6 to 8 weeks following the completion of initial assessments. The ESP will provide ongoing support to schools throughout implementation of the programme.
Who Provided (Intervention providers/implementers)	The basic concepts and design of Maths-Whizz were originally developed by graphic designer and author of interactive maths books, Ron van der Meer, and co-author and teacher Bob Gardner. Since then, it has continued to be adapted and enhanced by an in-house research and product development team at Whizz Education including STEM graduates from Oxford, Harvard and Imperial College London. Whizz Education will provide access to schools, teachers, pupils and parents to their respective areas of the Maths-Whizz platform. For each class, teachers can either allow the Tutor to determine the topics to be covered or they can choose Topic Focus to direct Maths-Whizz to directly support the topic they are teaching in class. When teachers choose Topic Focus, Maths-Whizz will set up differentiated lessons for each pupil in the class. Teachers can access assessments and progress reports to change course for each pupil depending on their changing needs and learning progress. Whizz Education also accesses the pupil, class and school assessments and progress reports to provide schools with targeted support and advise to optimise implementation.
Who Recipients	The Maths-Whizz virtual intelligent tutor provides personalised tutoring to pupils from Reception to Year 8. The Tutor can be

	implemented to the whole class or to a subset of pupils, but access to the platform is on an individual basis. For the purpose of this efficacy trial, Maths-Whizz will be available to the whole participating class of pupils in Years 2, 3, 4 and 5. While there are no specific exclusion criteria, pupils with severe visual disabilities will be unable to engage with the Tutor. Whizz Education, the delivery partner, will be recruiting schools to participate in the trial. The eligibility criteria for schools are provided in the 'Participants' section of the Impact Evaluation chapter.
How (Mode of delivery), when and how much	For this efficacy trial, pupils will be provided access to the Maths-Whizz intelligent tutor throughout the 2024/25 academic year, i.e., starting in September 2024 till July 2025. Schools will have the opportunity to discuss with Whizz Education the option of providing access to Maths-Whizz to pupils beyond the trial period. Pupils will engage in individualised learning through the Maths-Whizz online intelligent tutor. Lessons are automated and differentiated for every individual pupil depending on their unique needs and pace of learning. • When a pupil first logs in they complete a low pressure, interactive and adaptive initial assessment to identify individual strengths, weaknesses, and gaps in knowledge. • Maths-Whizz then provides interactive and engaging lessons, exercises, and tests in a sequence appropriate to each individual pupil's ability, constantly adapting at their own pace of learning with a focus on plugging knowledge gaps to accelerate pupil learning outcomes, reduce maths anxiety and build confidence. • All learning objectives are presented in three parts: (a) an animated lesson/tutorial that introduces the new concept or method; (b) an interactive exercise to determine whether the learner has internalised the concept by answering a set of questions and receiving help as needed; and (c) a brief test to establish the learner's level in the topic. Pupils only progress once they have passed the test. • As pupils complete learning objectives, their progress is available for teachers to monitor in a central reporting platform. The reporting platform empowers teachers, providing actionable information to support lesson planning for teachers, and further pupil intervention, including a wide range of usage and progress metrics. The programme consists of engaging pupils with the Maths-Whizz Tutor for about an hour each week. This can be divided into several sessions, such as three times a week (e.g., 3 x 20 minute sessions per week), and is provided in addition to

	regular maths lessons. Maths-Whizz may also be deployed within the school's homework policy. Details of how each school will implement Maths-Whizz will be determined during the implementation planning stage. Whizz Education will share regular reports of platform usage data for each pupil in the intervention year group with NFER. This will be used to calculate dosage of the intervention received by each pupil and its impact on their maths outcomes.
Where (Types of locations/necessary infrastructure)	The trial is open to primary schools across all regions of England. Where Maths-Whizz will be offered to pupils in school, schools are required to have sufficient number of devices such as laptops and/or tablets and sufficient wi-fi capacity. The minimum IT requirements for Maths-Whizz are detailed below: • Chrome 70+/MS Edge 90+/Safari 12+/Firefox 78+/Samsung Internet 11+ • Minimum 7.9" screen (recommended for usability reasons) • WebGL-enabled graphics card and browser • 2GB RAM • 0.5 Mbps internet connection per concurrent pupil
Tailoring (Adaptation of the intervention)	• The implementation of Maths-Whizz is tailored to each school's needs and goals during the Implementation Planning process. • The Maths-Whizz intelligent tutor adapts to the individual needs and pace of learning of each pupil. • The tutor maps out a tailored learning journey for each pupil based on the gaps identified in the initial assessment, which is interactive and adaptive. • Maths-Whizz calculates an overall 'Maths Age' for each pupil, which is indicative of the pupil's ability relative to the expected level of an average pupil of their age. • Teachers can access assessments and reports to identify further gaps and more targeted intervention where required. • Even when teachers choose to use Topic Focus to direct Maths-Whizz to align with topics being taught in class, Maths-Whizz will set up differentiated lessons for each pupil. • The programme can be offered to pupils in school or as part of their home learning, in which case pupils will access the Maths-Whizz platform at home.
Strategies to support implementation	• Whizz Education will develop a bespoke implementation plan with the school leadership and maths lead of each school that takes into account each school's needs and goals.

	• ESPs will provide continued implementation support to schools. ESPs monitor key school metrics on a weekly basis to identify schools that require additional support. • Schools will nominate a Maths Ambassador who will support the effective implementation of the programme in schools and communicate with the Whizz ESP. • Teachers will be able to access the initial assessment and learning data of all pupils to determine each child's needs and progress. • Parents will also be able to engage with their child's learning by monitoring their needs and progress. • ESPs organise termly check-ins with all schools to review and adjust their implementation plans, if required. • Whizz Education will organise an annual review meeting to review progress against the implementation plan and share key insights.

Logic Model

The logic model for Maths-Whizz is shown in Figure 1. It outlines the target population of the intervention and the activities, outputs, short-term and intermediate outcomes that lead to the outcomes both immediately after the intervention and longer term.

The primary causal pathway for the trial (shown by the yellow arrow in Figure 1) is that pupils' use of the Maths-Whizz tutoring platform for 45-60 minutes/week throughout one academic year (2024-2025) is expected to result in improved attainment, enjoyment and self-efficacy in maths, and improved engagement with maths teaching. The first three outcomes will be measured by the impact evaluation, whereas the fourth outcome, engagement with maths teaching, will be covered in the IPE. The Maths-Whizz tutoring platform embeds several mechanisms which are expected to improve pupil outcomes, these are outlined in Pupil Outputs and Short-Term Outcomes. As these mechanisms have not yet been tested in the context of Maths-Whizz, they are a key area to explore through the IPE.

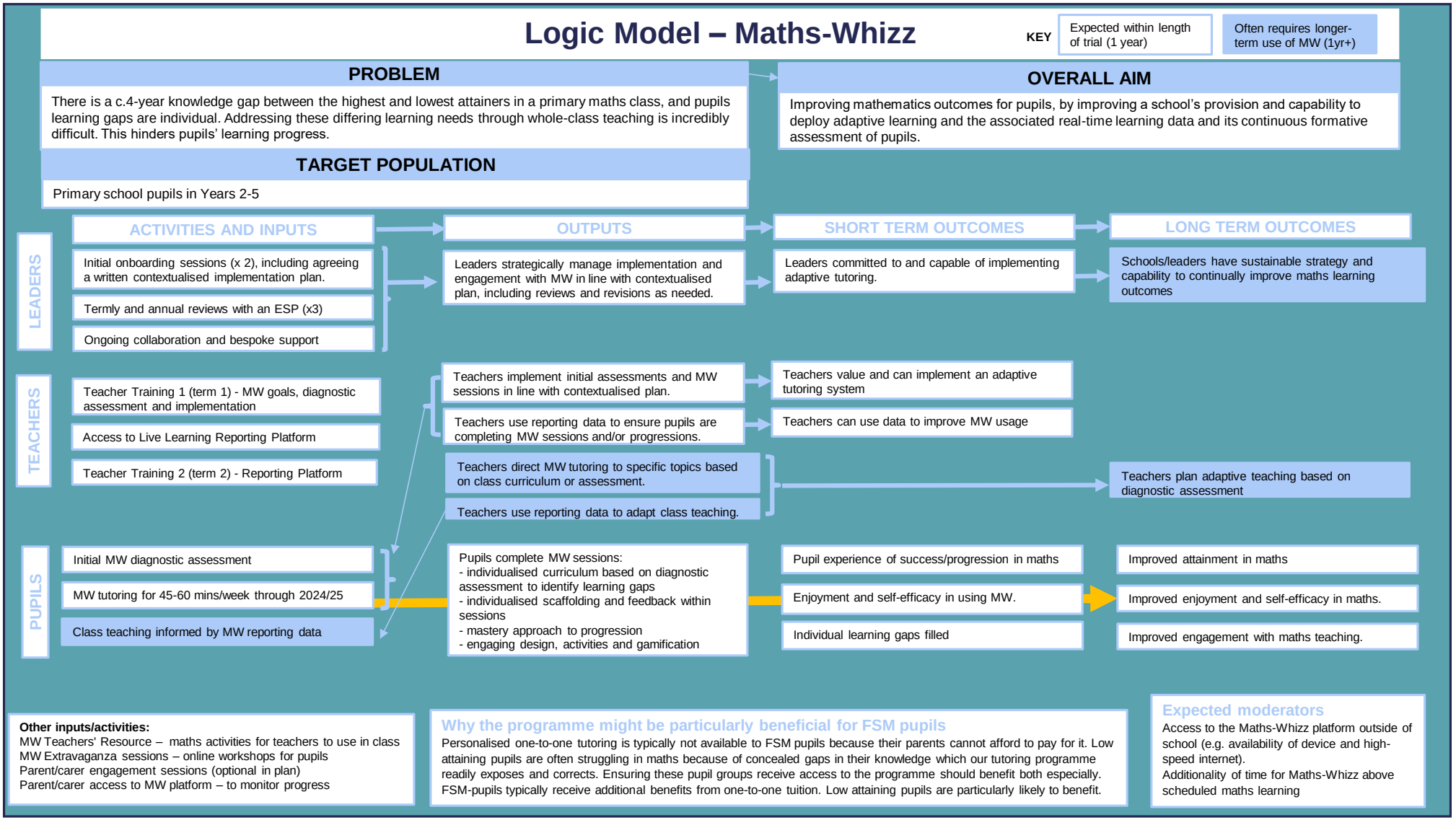
The expectation of the programme is that Maths-Whizz is used in addition to timetabled maths lessons, rather than a replacement for these lessons. However, Maths-Whizz may replace some maths learning time outside timetabled lessons, for example maths homework, pre-teaching or pre/post-school learning. We expect pupil impact will be moderated by whether Maths-Whizz use constitutes *additional* maths learning time or *replaces* other forms of maths learning time and this will be assessed in the IPE. We also expect that where a school's implementation model includes pupil use of Maths-Whizz at home, impact will be moderated at pupil-level by digital access at home (e.g. availability of a suitable device, and high-speed internet). We will collect preliminary data to explore the prevalence of this moderator, by collecting teachers' perceptions of barriers for pupils in using Maths-Whizz at home, and by monitoring any differences in usage and progression for (1) schools implementing home use (2) FSM pupils compared with non-FSM pupils, given likely differences in digital access.

Maths-Whizz provides inputs for school leaders and teachers (Activities and Inputs for Leaders and Teachers), which are captured separately in the logic model. As sustained pupil use of Maths-Whizz is a key mediator of impact, initial leader/teacher inputs are designed to support practical implementation, and develop buy-in/engagement from staff, to facilitate sustained pupil use. The support outlined in the logic model assumes relative stability of staffing, however responsive support and termly reviews may compensate to some extent for changes in staffing. Currently, Whizz provides extensive support for leaders and teachers, so it is important to understand what forms of support works, for which school staff, in what contexts. The IPE will explore the perceived impact of different aspects of support. It will also explore whether high levels of engagement with Maths-Whizz support is associated with sustained use of Maths-Whizz. This will help to inform the prioritisation of school support for potential scale-up.

Core use of Maths-Whizz (white boxes in Figure 1) comprises implementing and monitoring sustained pupil use of the Maths-Whizz programme, which is expected to improve pupil outcomes, and to facilitate staff commitment to adaptive tutoring. Whizz expect that the trial period (one academic year) will be sufficient for schools to implement core use and achieve those outcomes.

Some uses and outcomes of Maths-Whizz are typically associated with longer time-frames than the one-year trial period (blue boxes in Figure 1). Enhanced teacher use of Maths-Whizz involves stronger integration of Maths-Whizz into maths teaching and learning (teacher direction of the Maths-Whizz curriculum, and/or using Maths-Whizz progress data to inform class teaching). These uses are expected to enhance pupil outcomes and increase teachers' use of adaptive teaching in maths. These longer-term uses and outcomes may be seen in some schools within the trial period and will be explored through the IPE.

Figure 1: Maths-Whizz Logic Model



Impact evaluation

Research questions

The primary research question is:

RQ1: What is the overall impact of Maths-Whizz on the maths outcomes of children in years two to five?

The secondary research questions are listed below:

RQ2: What is the impact of Maths-Whizz on each individual year-group?

RQ3: What is the impact of Maths-Whizz on disadvantaged children as indicated by their FSM eligibility status?

RQ4: Does the impact of Maths-Whizz vary for different genders?

RQ5: What is the impact of Maths-Whizz on children with low prior learning in Maths?

RQ6: What is the impact of Maths-Whizz on children's motivation towards maths learning?

RQ7: What is the relationship between time spent (on tutor or practice modes) using Maths-Whizz (dosage) and improvement in maths outcomes?

RQ8: What is the relationship between Maths Whizz estimated maths-age and maths outcomes?

Design

Table 2: Trial design

Trial design, including number of arms		Two-arm, interleaved, three-level (pupil, class, school) cluster randomised efficacy trial
Unit of randomisation		School
Stratification variables (if applicable)		n/a
Primary outcome	variable	Maths attainment
	measure (instrument, scale, source)	Renaissance Star Maths Assessments scaled scores Unified Scaled Score ranging from 600 - 1400 (https://www.renlearn.co.uk/)
Secondary outcome(s)	variable(s)	Motivation towards maths learning
	measure(s) (instrument, scale, source)	Maths and Me survey consists of two scales: Mathematical self-perceptions Enjoyment of mathematics (Jill L. Adelson & D. Betsy McCoach (2011) Development and Psychometric Properties of the Math and Me Survey, Measurement and Evaluation in Counseling and Development, 44:4, 225-247, DOI: 10.1177/0748175611418522)

Baseline for primary outcome	**variable**	Maths attainment
	measure (instrument, scale, source)	Renaissance Star Maths Assessments scaled scores Unified Scaled Score ranging from 600 – 1400 (https://www.renlearn.co.uk/)
Baseline for secondary outcome	**variable**	Motivation towards maths learning
	measure (instrument, scale, source)	Maths and Me survey consists of two scales: Mathematical self-perceptions Enjoyment of mathematics (Jill L. Adelson & D. Betsy McCoach (2011) Development and Psychometric Properties of the Math and Me Survey, Measurement and Evaluation in Counseling and Development, 44:4, 225-247, DOI: 10.1177/0748175611418522)

This is an efficacy trial to evaluate the impact of the Maths-Whizz virtual intelligent tutoring programme over one academic year. Randomisation will be at the school-level; we will randomise 64 schools in a 1:1 ratio to two arms: schools in arm 1 will deliver Maths-Whizz to pupils in Years 2 and 4 and schools in arm 2 will deliver Maths-Whizz to pupils in Years 3 and 5. Pupils in year groups not receiving the programme in each arm will serve as controls for intervention pupils in the other arm. The programme will be delivered to pupils using a whole class approach. Interested schools will complete an Expression of Interest (EoI) and schools identified to be eligible will sign a Memorandum of Understanding (MoU) to confirm their participation. Pupils in participating classes in all four year-groups (Years 2, 3, 4 and 5) will complete the baseline assessment prior to randomisation. The primary outcome for this trial is maths attainment as measured by the Renaissance Star Maths scaled scores. The secondary outcome for this trial is motivation towards maths learning measured by the Maths and Me survey. This validated survey consists of two scales: mathematical self-perceptions and enjoyment of mathematics.

Randomisation

Schools will be randomised into two arms in a 1:1 ratio.

- In arm 1, pupils in Years 2 and 4 will receive the programme. Pupils in Years 3 and 5 will not receive the programme.
- In arm 2, pupils in Years 3 and 5 will receive the programme. Pupils in Years 2 and 4 will not receive the programme.

Pupils not receiving the programme in each arm will serve as the comparison group for pupils receiving the programme in the other arm.

There will be no stratification when randomising. Randomisation will be performed by NFER statisticians using R Code, which will be stored for reproducibility and transparency. The syntax used for randomisation will be appended to the Statistical Analysis Plan (SAP) and report. Test administrators will be blind to group allocation, but analysts will not.

Participants

Maths-Whizz will be delivered to the whole class rather than to individuals or subsets of pupils. Pupils in Years 2, 3, 4 and 5 are eligible to receive the programme. The teachers for each class will be responsible for managing the delivery of the programme and will participate in the Implementation and Process Evaluation (IPE), alongside school leaders and Maths Ambassadors.

Whizz Education will be responsible for the recruitment of 64 schools to this trial. They will do so through a combination of contacting schools on their mailing list and reaching out to new schools not on their contact list. If required, EEF and NFER will support their recruitment efforts by promoting the trial through their newsletters and social media channels. Whizz Education will offer webinars during the recruitment period to provide information on the trial to schools interested in participating. Interested schools will complete an online EoI, which will help Whizz Education ascertain schools' eligibility to participate in the trial. School eligibility criteria are set out below. Whizz Education will follow up with eligible schools via a phone call to confirm eligibility and clarify details such as the number of classes in each year group that will participate in the trial. Eligible schools will then be sent the school information sheet and Memorandum of Understanding (MoU). Schools will sign up to the trial by the headteacher signing on the MoU and providing the name of a key project contact who will act as the coordinator of the trial in the school. Once a school has signed up, Whizz Education will share the school name and details of the headteacher and key project contact with NFER. NFER will get in touch with the key project contact to initiate data collection and other trial-related activities.

School eligibility

Schools are eligible for the trial if they meet the following criteria:

- Be an Infant/Primary school with children in Year 2, 3, 4 and 5 as of 1 September 2024.
- Not have mixed year maths teaching even if the school has mixed year groups (i.e. children who are from more than one-year group within the same class) in any of Years 2, 3, 4 and 5 in the 2024-25 academic year.
- Not have implemented Maths-Whizz or any Whizz associated services in the 2023-24 academic year.
- Not be participating in another EEF-funded trial in the 2024-25 academic year.
- Have a suitable level of IT provision including devices such as tablets/laptops/desktops and sufficient wi-fi capacity (0.5 Mbps internet connection per concurrent pupil) (see Maths-Whizz minimum requirements).
- Agree to contribute **£500** towards the cost of the programme (usual cost £1875 per school plus £25 per pupil per year).

Pupil eligibility

There are no pupil eligibility criteria as the Maths-Whizz programme will be accessed by the whole class. Pupils with severe visual impairments will not be able to access the programme.

Sample size calculations

Table 3: Sample size calculations

		OVERALL	FSM
Minimum Detectable Effect Size (MDES)		0.11	0.13
Pre-test/ post-test correlations	Level 1 (pupil – KS1)	0.376	0.376
	Level 1 (pupil – KS2)	0.694	0.694
	Level 3 (school – KS1)	0.263	0.263
	Level 3 (school – KS2)	0.502	0.502
Level 3 (school – KS2)	0.037	0.037	0.019
	Level 2 (year group – KS2)	0.073	0.073
	Level 3 (school – KS1)	0.010	0.010
Alpha		0.05	0.05
Power		0.8	0.8
One-sided or two-sided?		Two-sided	Two-sided
Average pupils per class		26.7	26.7
Average classes per year group		1.4	1.4
Proportion of pupils eligible for FSM within last six years			24.6%
Number of pupils	Intervention	4,336	1,067
	Control	4,336	1,067
	Total	8,672	2,133
Number of classes	Intervention	162	162
	Control	162	162
	Total	**325**	**625**
Number of schools	**Total**	**58**	**58**

The assumed parameters for the sample size calculations are set out in the table above. The assumptions about average class size, the average number of classes per year-group and the proportion of pupils who have ever been eligible for FSM are based on the latest data from DfE (DfE, 2023). ICCs and pre-post correlations have been taken from the latest EEF guidance on power calculations (Singh et al., 2023), taking the median values based on NPD data between 2012 and 2019. We were not able to find any data looking at how the ICC for primary maths would divide if separated into year-group and school levels. Our assumption is that the vast majority of variance would be accounted for at the year-group level as these children will be being taught the same curriculum usually by the same teacher. In light of the absence of strong evidence supporting this assumption we took a relatively conservative position and split the ICC with one third at the school level and the remainder at year-group level. We also assume a pupil level attrition of 15% in line with our experience in similar trials. Sample size

has been estimated using the PowerUpR (Bulus et al., 2021) package within R. The syntax for the power calculations can be found in Appendix 1.

While the primary outcome for this evaluation is Renaissance Star Maths score amongst all pupils, the sample size was specified so that the evaluation would be powered for a minimum detectable effect size (MDES) of 0.2 for the analyses of individual year-groups, which, in practice, is the limiting factor. By powering the trial for the individual year-groups analysis, the trial is powered for all pupils by default. This specification gives quite small MDES' of 0.11 for the primary research question and 0.13 for the FSM analysis and requires 58 schools to achieve.

The school and pupil numbers given in Table 2 are those anticipated at the point of randomisation. We have not included any school-level attrition in these sample size calculations. Therefore, we have budgeted to over recruit by 10% to cover any school-level attrition and are intending to recruit 64 schools overall.

Outcome measures

Primary outcome

We will use Renaissance Star Maths assessments[2] as the primary outcome measure. These tests are listed in the EEF attainment measures database[3] and achieve the maximum possible ratings for all criteria.

The Star Maths assessments are computer-adaptive, adapting to the individual to provide an assessment that identifies gaps in learning from the entirety of the curriculum, independent of their current year group (Renaissance Learning, 2014). The assessment (which draws on an item bank of 2,000 items) comprises 24 items and can be completed by pupils (unaided by teachers) in 20 – 30 minutes. Star assessments are scored immediately upon completion and produce a 'Unified Scaled Score' ranging from 600 – 1400 across all year groups from Year 1 to Year 13. They also include a UK norm-referenced standardised score which takes account of the pupil's age comparing attainment with a pupil's peers nationally.

In an equating study conducted by NFER (Sewell *et al.*, 2007), Star Maths and Progress in Mathematics (PiM) tests were found to have correlations between 0.58 – 0.75. Star Maths also strongly correlated (correlation of 0.84) with the English National Curriculum equivalents given in PiM. Star Maths assessments are also accurate predictors of Key Stage 2 (KS2) SATs outcomes (*Relating results from Renaissance Star Reading™ and Renaissance Star Maths™ to the Key Stage 2 Standardised Attainment Tests (SATs)*, 2017).

In addition to their very high quality, these tests have several advantages which make them attractive for use in this evaluation:

1. They are vertically linked, so children from all year-groups can be placed on the same scale. This allows us to combine data from all four years giving very high power for both the primary research question and the FSM subgroup analysis.
2. They are completed online so the results can be easily accessed without physically moving scripts.
3. They are automatically marked thus reducing burden on teachers and making test results available immediately.
4. The infrastructure surrounding them has been developed with evaluators in mind. Tests can be administered at school level whilst the results are only available to the evaluators until they choose to release them to schools.

Using these tests for both baseline and endline assessment thus avoids relying on baseline KS1 or Early Years national assessments which are likely to have been disrupted by the pandemic for older children and would have considerably lower correlations with endline.

The baseline Star Maths assessment will be administered prior to randomisation by the schools via an online assessment. The assessment will be administered to all pupils in participating classes in Years 1, 2, 3 and 4 in summer 2024. By contrast, the endline RL Star Maths assessment will be overseen by independent NFER Test Administrators and will be

[2] https://www.renlearn.co.uk/
[3] https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluation/eef-outcome-measures-and-databases/attainment-measures-database/am-database

administered to all pupils in participating classes when they are in Years 2, 3, 4 and 5 in summer 2025.

Secondary outcomes

Secondary research questions will also explore key constructs in the logic model for Maths-Whizz relating to short-term and longer-term outcomes: enjoyment and self-efficacy (see Figure 1). These outcomes will be explored using the validated Maths and Me survey (Adelson and McCoach, 2011) that consists of two scales: mathematical self-perceptions and enjoyment of mathematics, which are measured by 18 items scored on a five-point Likert scale. This has been developed in the USA and is suitable for use with relatively young children (age 8-12 years). In the Spectrum database[4], it receives a two-star rating for psychometrics and three-star for implementation, but these scores may be too low as the measure has been developed further with adaptions for use in Spanish and Turkish (Cansiz and Tabak, 2018; Paz-Albo and Hervás-Escobar, 2023). This instrument is particularly attractive for this evaluation as it was deliberately designed to use simple language and it has relatively few items making it quick to administer. We propose to explore concepts such as pupil responsiveness/engagement and maths anxiety in our IPE.

The baseline Math and Me survey will be adapted by the study team to make it suitable for administration in UK schools (this simply involves replacing the word 'math' with 'maths' throughout). The survey will be administered by schools and will be completed online as soon as possible after the baseline Star Maths assessment is undertaken. Although the survey uses simple language, we think that it will not be suitable for Year 1 pupils. Therefore, at baseline, the Maths and Me survey will be administered to all pupils in participating classes in Years 2, 3 and 4 in summer 2024. Collection of the endline Math and Me survey will be overseen by NFER Independent Test Administrators alongside the collection of the endline primary outcome measure. At endpoint, Maths and Me will be administered to all pupils in participating classes when they are in Years 3, 4 and 5 in summer 2025.

Baseline measures

Both the primary and secondary outcomes will be measured at baseline as well as endpoint using the same instruments as outlined above.

Compliance

Compliance for this trial will be measured using data from the **Maths-Whizz platform and will be assessed at the class-level.** A class-level compliance measure is more suitable to this trial as it would help us understand whether Maths-Whizz would, on average, benefit *all* pupils if a school were to implement it as expected. In this scenario, there will be individual pupils who may not access the Maths-Whizz platform as expected. By taking a class-level approach, these pupils will not be excluded from the analysis, thus allowing us to assess the impact of compliance for *all* pupils within the school. A pupil-level compliance measure, on the other hand, would only help us understand whether or not Maths-Whizz would benefit pupils who use the platform as intended – leaving open the possibility that challenges facing pupils who don't engage as intended might outweigh the benefits for those who do. While a pupil-level compliance measure might be useful in itself, the class-level compliance measure is likely to

[4] https://educationendowmentfoundation.org.uk/spectrum-essential-skills-and-non-academic-outcomes/spectrum-database

be of greater relevance and use to schools in ascertaining the impact of the programme on *all* pupils when implementation is carried out as intended.

For this trial, a class will be deemed compliant if:

Pupils in the class spent at least 30 mins **Tutor** usage per week, on average, in the total period considered (termtime only/excluding the holidays)

OR

Pupils in the class attempted 3 **Tutor** lessons per week (pass, static or fail), on average, in the total period considered (termtime only/excluding the holidays).

We will explore the linear effect of this compliance indicator on the primary outcome measure. A complier average causal effect (CACE) estimate will be obtained using instrumental variable modelling (details will be provided in the SAP) and will be obtained for the primary outcome measure if the finding for the ITT analysis for the primary outcome measure is not statistically significant.

Analysis

Primary and secondary analyses

The main analysis for this trial will be intention-to-treat and will follow the EEF analysis guidance[5]. The primary outcome measure is the Star Maths assessment scaled score, which will be the dependent variable in a multi-level regression model with an independent indicator variable representing the status of the intervention for that student (1 = intervention, 0 = control). Students' baseline Star Maths assessments will be used as a covariate in this multi-level model that models both the class-level and the school-level clustering as random effects. Effect size will be calculated using Hedge's g, using total variance from a model without covariates. 95% confidence intervals will be computed for the effect size.

Similar models will be run for the secondary analysis. For this analysis, the scores from the two subscales of the Math and Me survey ('Mathematical self-perceptions' and 'enjoyment of mathematics') will be the dependent variables in two separate multi-level models.

All analysis will be carried out within R, with modelling performed using the nlme and lme4 packages.

Subgroup and exploratory analyses

While the primary research question assesses the impact of Maths-Whizz on all pupils in the trial, the secondary research question will evaluate the impact of Maths-Whizz on the maths outcomes for pupils:

(1) In each individual year-group (Years 2, 3, 4 and 5) to explore any differences in impact by year-group. This trial is powered for the individual year-group analysis.

(2) In alignment with EEF's focus on the educational attainment of pupils from disadvantaged backgrounds, we will assess the impact of Maths-Whizz on FSM-eligible pupils' maths outcomes. A binary indicator that shows whether a child has been

[5] https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/evaluation-design/EEF-Analysis-Guidance-Website-Version-2022.14.11.pdf?v=1709155078

eligible for free school meals within the last six years (EVERFSM_6_P from the National Pupil Database) will be used as the identifier for disadvantage.

(3) Additionally, we will also explore whether the impact of Maths-Whizz on maths outcomes varies by gender (male and female). In the UK, at the age of 11, boys are four per centage points more likely than girls to achieve the expected standard in maths ((Borra, Lacovou and Sevilla, 2021)). We will explore whether an intelligent tutoring platform like Maths-Whizz has the potential to close this so-called gender gap in maths education.

(4) Prior attainment significantly affects maths outcomes of pupils as it not only influences pupils' ability to learn and consolidate new concepts but also affects their confidence in their mathematical abilities and contributes to higher levels of maths anxiety. We will explore whether Maths-Whizz can improve maths outcomes for pupils with low prior attainment. In the absence of Key Stage assessment data, we will consider using either the initial assessment scores or the baseline Star Maths score to define low prior attainment.

These analyses will also use a multi-level model as outlined for the main analysis. Each of these subgroup analyses will proceed by first assessing the effect of the intervention in that subgroup and then investigating a differential effect of the intervention using an interaction term between the subgroup indicator and the treatment indicator added.

The impact of dosage of the programme on pupils' maths attainment will be explored by analysing usage data (time spent by pupils on the tutor mode or time spent by pupils on practice mode) from the Maths-Whizz platform. The Maths-Whizz tutor prepares a differentiated learning journey for each pupil based on each pupil's Maths Age. We will explore the relationship between Maths Age and maths outcomes as measured by the Renaissance Star Maths scaled scores. Details of these analyses will be provided in the SAP.

A Complier Average Causal Effect (CACE) analysis will be carried out using the compliance measure outlined above (see compliance section).

Missing data analysis

Individual child-level attrition is unlikely to be a significant issue in this trial due to the class-wide implementation of Maths-Whizz. Likewise, school-level measurement attrition is anticipated to be relatively low within this trial. Despite this, there is likely to be some degree of missing data for the primary outcome at endline. Where pupils are unavailable for testing, the reason for this will be established where possible and described in the final report. Additionally, variables that are found predictive of missingness in the primary outcome will be added to the primary analysis model to explore their impact on the estimated effect of the intervention. Further missing data analysis will be undertaken if needed as per the EEF analysis guidance[6].

Full analysis will be pre-specified in a detailed SAP.

[6] https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/evaluation-design/EEF-Analysis-Guidance-Website-Version-2022.14.11.pdf?v=1709155078

Longitudinal follow-ups

No longitudinal follow-up is planned.

Implementation and process evaluation

Research questions

Our IPE follows EEF's updated IPE guidance and complements our impact evaluation by exploring *how* and *in what circumstances* Maths-Whizz impacts teachers and pupils.

As this is an efficacy trial, a key focus will be implementation fidelity. Given the complexity of the Theory of Change (ToC), we will identify how key implementation characteristics contribute to key outcomes to refine the ToC and inform potential scale-up. We will compare FSM-eligible pupils with non-eligible pupils to describe any differences in implementation, experience and perceived outcomes for FSM-pupils.

Considering how Maths-Whizz is implemented, our IPE will test the logic model's hypothesis about the leader/teacher support (activities, inputs and outputs) needed to facilitate weekly implementation of Maths-Whizz tutoring (RQ1). We will identify which supports listed in the logic model are perceived as most important in different school/teacher contexts, any other key support needed for implementation in different contexts (RQ2). We will also calculate the *cost* of delivery (RQ7). Considering perceived impact, we will refine the pupil outputs and outcomes in the logic model by describing teachers' and pupils' perceptions of impact and the key features and mechanisms which facilitate this impact (RQ3). Considering variation in approaches to using Maths-Whizz, we will describe the extent of integrated use of Maths-Whizz in maths teaching (shown as blue boxes in the logic model) and the perceived facilitators of this (RQ4). We will identify key moderators and mediators (RQ5) using both platform data and teacher perceptions, and we will refine the logic model by identifying which approaches to using Maths-Whizz are associated with strong outcomes (RQ6).

To contextualise our impact evaluation's comparison of the intervention and control groups, we will describe 'business as usual' for both arms, monitoring this during the trial for the control group, and allowing us to assess programme differentiation for the intervention group.

RQ1: To what extent was fidelity of implementation achieved? [Fidelity, Quality, Adaptation]

1.1 Did school leaders/teachers attend the relevant onboarding and reviews?
1.2 Was Maths-Whizz used for at least 30 minutes per week for each class?
1.3 Was Maths-Whizz used for at least 30 minutes per week for each pupil?
1.4 How did implementation vary across and within schools, including the additionality of Math-Whizz time and planned home usage?
1.5 Was there any difference in implementation in schools with different proportions of FSM pupils?
1.6 Were there any differences in Maths-Whizz use between FSM and non-FSM pupils?
1.7 Where fidelity was not achieved, why, and what adaptations were made?

Our impact evaluation will assess the impact of dosage on maths outcomes.

RQ2: What are the facilitators and barriers for implementing Maths-Whizz? [Context, Responsiveness]

We identified three key implementation factors which informed our research questions: quality of IT infrastructure; teacher/leader engagement and efficacy (supported by Maths-Whizz implementation inputs); and how Maths-Whizz affects workload (EEF, 2019; DfE, 2022a).

2.1 What facilitated and hindered schools in using Maths-Whizz? (e.g., quality of IT infrastructure, devices, spaces, timetabling, teacher confidence, home use)
2.2 Were leaders, teachers, pupils and parents engaged? What facilitated and hindered this (e.g., Maths-Whizz support)?
2.3 Were any barriers/facilitators experienced specifically or differentially by FSM pupils?
2.4 How has Maths-Whizz affected teacher workload?
2.5 Overall, what worked well and less well, why?

RQ3: What are the perceived outcomes of Maths-Whizz? [Perceived impact]

3.1 What are perceived outcomes for pupils (e.g. attainment, motivation, enjoyment, confidence, self-efficacy, maths anxiety)?
3.2 What are perceived outcomes for teachers and schools (e.g. digital literacy, formative assessment, adaptive teaching, workload)?
3.3 Are there any unintended consequences for pupils, teachers or schools?

RQ4: What are mediators and moderators for Maths-Whizz outcomes? [Mediators, Moderators]

We identified three mechanisms which may contribute to improved maths outcomes. First, learning within Maths-Whizz may directly improve maths outcomes. Second, using Maths-Whizz may improve pupils' engagement with maths teaching, mediating improved maths outcomes (e.g., filling learning gaps improves pupils' access to class teaching). Third, changes in teacher practice (e.g., using Maths-Whizz data to tailor class teaching) may mediate improved maths outcomes.

4.1 To what extent are perceived pupil engagement with maths teaching and changes in teacher practice associated with enhanced Maths-Whizz outcomes?
4.2 Do teachers perceive any differences in Maths-Whizz usage or outcomes for different pupils?
4.3 Were there any differences in perceived outcomes for FSM pupils?

Our impact evaluation will assess the impact of Maths-Whizz for different pupil characteristics (e.g. gender, levels of prior attainment, FSM eligibility).

RQ5: What is 'business as usual', and to what extent was Maths-Whizz different? [Programme differentiation, Monitoring of control groups]

We expect considerable variation in the extent of differentiation between Maths-Whizz and business as usual. As relevant factors (e.g., digital literacy, assessment practices) vary significantly at teacher-level, we will collect teacher-level data at baseline and endpoint.

5.1 Prior to implementation, how did teachers use formative assessment, pupil data, EdTech and supplemental support in maths teaching?

5.2 What activities (if any) did Maths-Whizz displace? Did it add to or displace maths learning time, including maths homework?

5.3 To what extent did schools engage with wider Maths-Whizz resources (e.g. curriculum materials, pupil workshops) and what was the perceived impact of this?

5.4 What was usual practice in control year-groups during the intervention (e.g., supplemental support, EdTech and formative assessment in maths)?

RQ6: What implementation characteristics improve maths outcomes for pupils?

To understand *how* Math-Whizz affects outcomes and refine the ToC for scale-up, we will explore how a small number of implementation characteristics (e.g. Maths-Whizz usage, additionality of Maths-Whizz time, home use, programme differentiation) relate to a specific outcome (e.g. progress in Maths-Whizz or attainment). The preferred implementation characteristics and outcome will be chosen in spring 2025, after collecting Early Implementation Data (see Research Methods below), in order to reflect variations in implementation within the sample of participating schools.

6.1 How do key aspects of implementation relate to a specific Maths-Whizz outcome?

RQ7: What is the cost per pupil per year of Maths-Whizz? [Cost]

7.1 What is the cost to schools per pupil per year to provide the Maths-Whizz Intelligent Tutoring programme?

Research methods

We propose an iterative IPE, with early findings influencing instrument design and sampling. This responsiveness is particularly useful at efficacy trial stage, where lines of enquiry often emerge as schools implement the programme.

The IPE lead will design the instruments and collect most of the IPE data, to ensure continuity across design, data collection and analysis. Case study visits will be undertaken by the IPE lead and one other trained qualitative researcher from NFER, who will be briefed on the research aims and instrument.

Set-up and baseline (Nov 2023 - Sep 2024)

IDEA workshop and set-up meetings to co-construct the TIDieR framework, understand the intervention and programme materials and refine the ToC to guide research questions.

Survey teachers (intervention and control) at baseline to explore business as usual, particularly teacher efficacy and practice (e.g., formative assessment, using pupil data, EdTech, supplemental maths support). Where possible, survey questions will be taken/adapted from validated teacher surveys or scales (e.g. TALIS).

Early implementation data (Nov 2024-Jan 2025)

Observe teacher training (session 1), one per Education Support Partner (ESP), including a trainer post-interview, focusing on fidelity of training, adaptations, and teacher responsiveness.

Review implementation plans for all schools to understand the range of implementation approaches (e.g., extent of home learning, intended frequency of delivery, aims for using Maths-Whizz), anticipated facilitators and barriers, and goals for using Maths-Whizz.

Review Maths-Whizz platform and records for pupils (usage, progression) and parents (engagement). This will include a school-level comparison of usage and progression for FSM pupils, compared with other pupils.

Conduct a **focus group** with Maths-Whizz ESPs, exploring school support, school-level variance in early implementation, and facilitators and barriers.

Case study visits (Feb–Mar 2025)

Visit six **case study** schools (c.10% sample,) chosen to allow for some variation in school characteristics and implementation approaches. The case study visits are intended to be exploratory, and from six school visits we do not expect to achieve saturation in terms of the range of practice and contextual characteristics. The in-depth exploration of case study schools will be complemented by the ESP focus group, platform data review and teacher survey which cover all schools.

Schools will be sampled purposively based on FSM-levels and implementation characteristics. If possible within the group of recruited schools, we will include 2-3 schools with a proportion of FSM-eligible pupils above the national average, to explore the logic model hypotheses about differential impact for FSM pupils and the relevance of digital access at home, and to understand specific barriers and facilitators in these contexts. If present within the sample, we will include 1-2 schools who are implementing more integrated/long-term approaches to using Maths-Whizz, such as directing the Maths-Whizz curriculum or using Maths-Whizz data to adapt class teaching. The ESP focus groups, implementation plans and platform data will be used to identify one other selection criterion for sampling schools. This criterion will reflect a difference in approaches to implementing Maths-Whizz which reflects the logic model or expected moderators, and schools will be sampled to cover the range of approaches. To maximise response from sampled schools, we will offer an

incentive of £150 for each case study school. If a school declines to participate in a case study visit, we will invite a similar school based on the sampling frame.

- Interviews with senior leader, Maths-Whizz ambassador and intervention teacher, focusing on their use of Maths-Whizz according to their role (training and support, use and integration, differentiation compared with business as usual, facilitators/barriers, perceived pupil engagement/outcomes, changes in cost/time)
- Focus group with Y4 or Y5 intervention pupils (experience of use, engagement and perceived outcomes). We chose Y4/5 over the younger year groups because we expect older pupils to be more fluent and reflective in describing their perceptions and experiences. We will randomly select one intervention class from the case study school for the focus group, so that pupils are familiar with each other. We aim to include 5-6 pupils in each focus group, to include pupils with different characteristics while keeping the group size manageable so that all pupils are heard. We expect focus groups for this age group to last a maximum of 45 minutes in order to sustain attention. Teachers will be asked to select pupils in advance, including at least one pupil of higher-, medium- and lower-attainment within the class context. We will ask teachers to construct a group of mixed gender, and, where possible, to include at least one FSM-eligible pupil.

Endpoint data (Jun 2025)

Review Maths-Whizz platform data/records (as above).

Survey school staff at endpoint. Where possible, survey questions will be taken/adapted from validated teacher surveys or scales (e.g. TALIS). All questions will be closed, with a mixture of single-response and multiple-response items.

- Intervention teachers: training, Maths-Whizz use and integration, changes in practice, perceived pupil engagement/outcomes, moderators and mediators and costs/time, also repeating baseline questions to capture changes in efficacy/practice (e.g., formative assessment, using pupil data, EdTech, supplemental maths support).
- Maths-Whizz Ambassadors and school leaders: onboarding/reviews, school-level implementation, changes in practice, perceived engagement/outcomes and costs.
- Control teachers: maths teaching practice during the intervention, and repeating baseline questions to explore changes and contamination.

Table 4: IPE methods overview

Research methods	Data collection methods	Participants/ data sources and sampling	RQs	Data analysis methods	Implementation/ logic model relevance											
					Context	Fidelity	Quality of delivery	Compliance/ dosage	Reach and responsiveness	Adaptation	Mediators incl. intermediary outcomes	Causal mechanisms	Moderators	Secondary outcomes	Usual practice	Cost data
IDEA workshop	TIDieR framework; logic model	EEF, developer, NFER team	1-7	Descriptive analysis	✓	✓		✓			✓	✓	✓	✓		✓
Baseline survey	Teacher survey	Intervention and control teachers (all teachers, n~300)	5	Statistical analysis	✓								✓		✓	
Onboarding observations	Structured observation and follow-up interview	MW Education Success Partners, schools (3 - one per partner)	1,3	Thematic analysis	✓	✓	✓	✓	✓	✓			✓		✓	
Review of school implementation plans	Desk review	School implementation plans (all schools, n=58)	1-4,6	Qualitative comparative analysis	✓	✓	✓	✓	✓	✓			✓		✓	
Interim MW platform data	Desk review	Pupils, teachers and parents from all schools (population)	1-4, 6	Statistical analysis	✓	✓	✓	✓			✓	✓	✓			

Focus group with Education Success Partners	Semi-structured online focus group	MW Education Success Partners (all, n~3)	1-4, 6	Thematic analysis	✓	✓	✓	✓		✓	✓	✓	✓	✓		
Case study visits	Interviews with school leader, MW Ambassador, intervention class teachers. Pupil focus group Y4 or Y5.	Six schools, sampled to cover a range of implementation characteristics and high/low FSM proportions. 5-6 pupils (varied attainment, FSM, and gender)	1-5, 7	Thematic analysis	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	
End-point survey	Online survey	All schools (n=58). All intervention teachers (-150), control teachers (~150), MW ambassadors (n=58), school leaders (n=58)	1-5, 7	Statistical analysis	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓
MW Platform data	MW platform data	Pupils, teachers and parents from all schools (population)	1-6	Statistical analysis		✓		✓	✓	✓	✓	✓	✓			

Analysis

Qualitative Data – training observations, ESP focus groups, case study interviews and pupil focus groups

Training observations will be analysed thematically, considering fidelity/adaptations, facilitators/barriers and teacher responsiveness.

Qualitative data from the ESP focus group and the six school case studies will be recorded, transcribed, and analysed thematically in MAXQDA. We will develop a deductive top-level coding frame based on the research questions (e.g. fidelity and adaptations, facilitators/barriers), and will code the data from each source inductively within that frame. The inductive coding will then be compared with the logic model (e.g. comparing perceived outcomes reported by teachers and pupils with the outcomes anticipated in the logic model).

For the case studies, we will use MAXQDA to combine the qualitative data with key implementation characteristics (e.g. extent of home-use, and average pupil usage). Case-oriented thematic analysis will provide a rich description of implementation for each school case (n=6), complementing the qualitative comparative analysis (described below) which covers all schools.

Qualitative comparative analysis

We will identify key implementation and contextual characteristics of schools (chosen after collection of Early Implementation Data) from the teacher survey data and school implementation plans and categorise these for each school. Characteristics will be selected if they (i) align with expected moderators or other aspects of the logic model, for example teachers using Maths-Whizz data to adapt teaching, or the planned proportion of home use and (ii) vary across the sample of schools. We will explore this data using qualitative comparative analysis (Cilesiz and Greckhamer, 2020), a qualitative technique which maps variation across cases (schools) and identifies the combinations of characteristics which are consistently associated with an outcome's occurrence. With a sample size of 58 schools, we anticipate that we can map up to four implementation and contextual characteristics. We will compare these characteristics with school-level outcome data (e.g. the presence/absence of high Maths-Whizz progression or attainment, defined with reference to the variation in the sample) to identify combinations of implementation and contextual characteristics which are consistently associated with the chosen outcome.

Quantitative data - surveys

We will summarise endpoint data with descriptive statistics, including frequencies and cross-tabulations (e.g. with FSM eligibility) and use linear regressions to relate key implementation characteristics, such as additionality of Maths-Whizz time, to outcomes. We will compare baseline and endpoint teacher data to explore changes in teacher practice, contextual changes such as maths professional development or intervention programmes, and any self-reported contamination (e.g. use of similar maths ed tech). As an initial exploratory analysis of potential pathways for the impact of Maths-Whizz, we will use cross-tabulations to check whether changes in (1) perceived pupil engagement with maths teaching outside Maths-Whizz provision and (2) teachers' classroom practice are associated with enhanced attainment outcomes.

Synthesis of IPE data

For each data source and participant group, we will summarise the findings for each relevant research question. We will systematically compare and contrast findings for each data source and participant group, for example comparing perceptions of barriers and facilitators across Maths-Whizz ESPs, leaders and teachers. We will highlight areas of congruence and difference in reporting. Where the sequence of activities allows, we will explore areas of difference directly with participant. For example, we will ask case study interviewees about the reasons for any differences between the intentions recorded in school implementation plans, and the implementation in practice (based on the review of platform data).

The Research Methods section above describes how findings from initial IPE activities will analysed to inform subsequent IPE activities, while the Analysis section describes how qualitative comparative analysis will collate data from different IPE activities to identify characteristics associated with positive outcomes.

Cost evaluation

In line with EEF's latest cost evaluation guidance, we will conduct a cost evaluation that robustly estimates the market cost per-pupil-per-school-year of participating in Maths-Whizz, incorporating both financial and time costs. Data collection for the cost evaluation will be incorporated into other planned IPE activities (below) to minimise burden for schools.

The main financial cost is expected to be the cost of the Maths-Whizz programme, which will be collected directly from Maths-Whizz. As the trial inclusion criteria specify that schools should already have the technological infrastructure needed to implement Maths-Whizz, we do not expect schools to incur additional costs related to technology. However, we will check this, and other financial costs, through the school leader survey and interview.

We will compare the technology requirements specified for Maths-Whizz with teachers' perceptions of the requirements. We will compare these requirements with TIMSS 2023 data on digital infrastructure (available from spring 2025) to estimate the proportion of schools in England who would meet these requirements.

Time costs will be estimated in comparison with business-as-usual by identifying time investments and savings in using Maths-Whizz, through the review of implementation plans, staff surveys and staff interviews. Time costs to be considered will include the time spent on training, set-up and planning, any additional time supervising Maths-Whizz (e.g. through a homework club) and time spent monitoring, reporting and reviewing. Time savings to be considered will include reduced teacher contact time (e.g. if another member of staff is supervising Maths-Whizz use within lesson time), reduced planning time for lessons, and reduced assessment/marking time.

We will review school implementation plans to explore the variation in the extent of home use of Maths-Whizz across schools, as we would expect home use to have different time costs and/or savings compared with school use. If there is significant variation in home use, we will undertake sensitivity analyses to estimate costs for different levels of home use.

To allow comparison with other trials, we will estimate a per-pupil-per-year cost for programme use across three years, by adding one-off costs (such as training) and annual ongoing costs (such as the financial subscription, and planning/teaching time).

Ethics and registration

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

This trial will be registered at the ISRCTN registry and the trial registration details will be updated in this protocol as soon as it becomes available. The trial registry will also be updated with outcomes at the end of the project.

Each participating school's headteacher will provide their agreement to participate in the trial by signing the Memorandum of Understanding (MoU) that outlines the responsibilities of all parties involved in the trial. NFER will share a parent letter and withdrawal form with schools to be sent to parents/carers of all pupils that schools intend to nominate for participation in the trial. Through the withdrawal form, parents/carers will have the opportunity to withdraw their child from the evaluation and associated data processing at any stage of the trial.

A separate opt-in agreement process will be used for the pupil focus groups and will only apply to those selected to participate. Since pupils participating in the focus groups are only 8 to 10-years-old, we cannot assume that all pupils will have the capacity to provide fully informed consent to participate. In addition, as the focus groups involve audio recordings, it is especially important to ensure that parents/carers have the option to specifically agree to their child participating in this evaluation activity. We will, therefore, provide parents/carers with a written information sheet which will contain full details about the focus group and what their child will be asked to do. Parents/carers will then be asked to provide written opt-in agreement for their child to be invited to participate in the focus group, by returning a consent form to the school, who will then pass this information on to the research team.

Pupil participation in the focus groups is voluntary, therefore even if a parent/carer has agreed for their child to participate, their child can still choose not to take part. Age-appropriate information about the focus groups will be provided to pupils at the same time as parents/carers receive information about the focus groups to allow them to discuss participation together. The researchers will also read this information to pupils at the beginning of the focus group to ensure pupils understand it and have the chance to ask any questions. If at any point a pupil decides that they would prefer not to participate, then they will be able to return to their class. Prior to beginning the focus group, the researchers will agree some ground rules for the group with the pupils and have a discussion with them about the types of scenarios in which we would need to break confidentiality, to ensure they fully understand what this means.

Data protection

All data gathered during the trial will be held in accordance with the data protection framework created by the Data Protection Act 2018 and the General Data Protection Regulation 2016/679 and will be treated in the strictest confidence by the NFER, Whizz Education and EEF. No individual or school will be identified in any report.

NFER and Whizz Education are independent data controllers for the evaluation and the programme, respectively, for the duration of this trial.

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

NFER and Whizz Education have signed a Sharing Agreement that will govern the collection and sharing of personal data during this trial. This agreement includes a description of the nature of the data being collected and how it will be shared, stored, protected and reported by each party. In addition, Whizz Education will provide a memorandum of understanding to schools, explaining the nature of the data being requested of schools, teachers and pupils, how it will be collected, and how it will be passed to and shared with NFER. Two separate Privacy Notices, one for schools and another one for parents, are available here.

For the purposes of the trial, Whizz Education will collect names, role and contact details of a key contact person and the person signing the MoU when schools are recruited. They will share these data with NFER, who will then contact the key project contact person at participating schools to initiate data collection for the trial. NFER will ask participating schools to share pupil data for all pupils in each participating class in Years 1, 2, 3 and 4 and will subsequently share these data only for pupils in the intervention year groups with Whizz Education. NFER will also collect data for class teachers in each participating class. This data will be used to administer baseline and endpoint surveys to teachers and will also be shared with Whizz Education to facilitate the coordination of training. Whizz Education will also share the names and contact details of the Education Support Partners (ESPs) with NFER, so that NFER can contact them to attend and observe a sample of training sessions and conduct a focus group with ESPs. All personal data will be shared via secure, password-protected data sharing portals.

NFER will also collect pupil data from schools including names, date of birth, Unique Pupil Number (UPN), gender, FSM eligibility status, current class name and class name in next academic year for all pupils in participating classes in year groups 1, 2, 3 and 4. The baseline and endpoint Renaissance Star Maths score and the Maths and Me survey responses will be collected for all these pupils. In addition to these data, for pupils in the intervention year groups, NFER will also receive reports of Maths-Whizz platform usage data. For all pupils in the trial, background data including gender and FSM eligibility status will be collected from the National Pupil Database (NPD). To obtain the information from the NPD, NFER will provide the Data Sharing Team at the DfE with the names of the pupils, their dates of birth and UPNs, allowing a match to NPD.

As part of the implementation and process evaluation (IPE), NFER will conduct online surveys of teachers, leaders and Maths Ambassadors, observations of training sessions, focus group with trainers (ESPs) and interviews with teachers. A small number of schools will also be invited to participate in pupil focus groups. All NFER staff visiting schools will have up-to-date DBS checks. All data gathered during interviews will be stored securely. No names of individuals will be used in any report arising from this work.

Within three months of the end of project, NFER will send school and pupil data to EEF's data archive partner. At this point, EEF's data archive partner will keep a copy of the data and EEF will become the Data Controller. NFER will retain personal data for one year after report publication in case there are any queries about the report. One year after the report publication (expected to be May 2026), all personal data will be securely deleted.

Personnel

Name	Organisation	Role and Responsibilities
Stephen Welbourne	NFER	Project Director – responsible for overall delivery of the trial
Aarti Sahasranaman	NFER	Trial Manager – day-to-day management of the trial and delivery of the trial design and point of contact for EEF
Katherine Aston	NFER	IPE Lead – design and delivery of the IPE
Gemma Schwendel	NFER	Statistician
Kathryn Hurd	NFER	Research Operations Lead - overall data collection and school communications strategy
Jishi Jose	NFER	Project Manager – day-to-day operations including preparation of recruitment documents, coordinating data collection and point of contact for schools participating in the trial
Richard Marett	Whizz Education	Project Director – responsible for strategic oversight
Sarah Hawkes	Whizz Education	Project Manager – day-to-day operations including preparation of recruitment documents, coordinating recruitment processes/strategy, project management, and point of contact for EEF/NFER
Emma Ringe	Whizz Education	Project Implementation Manager – overall responsibility for recruitment of schools and implementation management during delivery period
Ray Douse	Whizz Education	Compliance Director – ensuring compliance with GDPR and other compliance measures

Risks

S. No.	Risk	Risk Assessment		Mitigation/Counter Measures/Contingencies
		Likelihood	Impact	
1	Insufficient schools recruited by Whizz Education (Target: 64)	Low	High	• NFER will input into recruitment material and work closely with Whizz Education. If required, our experienced operations team can assist with recruitment through a separate grant agreement. • Regular recruitment updates from Whizz Education and monthly check-in meetings with Whizz Education and EEF. • Decide and monitor pre-agreed recruitment targets to identify any unfavourable trends early on to act quickly. • The design envisages that both arms will receive the intervention in different year-groups, which will make this trial more attractive to schools. • Review sample size calculations if required and change parameters for power of the trial.
2	School attrition from trial and primary analysis	Low	Moderate	• Sign up to the trial via Memorandum of Understanding with clear identification of requirements. • Clear initial and ongoing communication with one key contact per school explaining principles, expectations, timelines and next steps. • Offer support webinar ahead of baseline data collection to allow practitioners to ask any questions. • Schools will only get randomised if they have completed baseline testing. • Both trial arms receive the intervention reducing the likelihood of post randomisation attrition. • NFER's independent Test Administrators, all ex-teachers, will serve to

				ensure very high follow-up rates at endpoint, visiting schools at a convenient time. • Provide incentive payments of £250 to all schools upon completion of endpoint testing. • Throughout the trial, Whizz Education will monitor platform usage to check that schools are using the intervention and attempt to re-engage any that are not.
3	Number of classes participating in the trial from each school exceeds assumptions impacting budget	Low	Moderate	• Assumption of number of classes in each year group based on national average for English schools. • Collect data on (and monitor) number of classes in each year group expected to participate in the trial during the EoI stage during conversations between schools and Whizz Education. • Allow schools to put forward up to three classes for each year group and consider participation of bigger schools on a case-by-case basis.
4	Schools may not complete baseline assessments within testing window	Low	High	• Provide schools with clear guidance on administering the baseline assessments. • Provide schools with a three week window to complete baseline assessments for all trial pupils and an additional week to complete any further testing if required. • Monitor completion of baseline assessments and follow-up with schools as required.
5	School doesn't have the necessary IT capacity for intervention delivery and outcome measurement	Moderate	Moderate	• Clear initial communication outlining IT requirements at EoI and MoU stage both in terms of WiFi connectivity and number and type of devices. • Verbal confirmation that schools meet this requirement obtained by Whizz ahead of sending MoU.

6	Intervention is not implemented as intended	Low	Moderate	• Clear information provided to schools explaining the principles of the trial and expectations. • IPE to monitor implementation fidelity
7	Difficulty in securing target response rates for IPE	Moderate	Moderate	• Communication with schools explaining research benefits • Ongoing reminders • Flexibility in timings of school visits • Close liaison with Whizz education to support IPE engagement • Online data collection where possible to minimise burden. • Provide school-level incentives of £150 for case study schools to encourage engagement with research and recognize time given to these activities.

Timeline

Table 5: Timeline

Dates	Activity	Staff responsible/ leading
Oct ’23 – Jan ‘24	**Project set-up** Complete project set-up, finalise recruitment documents and privacy notices, finalise grant agreement	Stephen Welbourne, Aarti Sahasranaman
Feb – May ‘24	**Recruitment and pupil data collection** Finalise data sharing agreement School recruitment by delivery partner Recruited school data from delivery partner in three tranches – one in April and two in May Collect pupil data for all Year 1 – 4 pupils (will be in years 2 – 5 in September) in participating classes	Kathryn Hurd, Jishi Jose
	Study protocol First draft of study protocol submitted in Mar ’24, revise based on feedback	Aarti Sahasranaman
Jun – Jul ‘24	**Baseline impact and IPE data collection** Set up Renaissance Star Maths accounts for schools and pupils Coordinate baseline assessments (Renaissance Star Maths, Maths and Me survey) in schools Baseline teacher survey	Kathryn Hurd, Jishi Jose, Katherine Aston (teacher survey)
	Randomisation Randomisation of schools in Jul ‘24	Gemma Schwendel
Aug ’24 – Jul ‘25	**Maths-Whizz training, delivery and data sharing (by Whizz Education)** Teacher training and onboarding in Aug – Sep ’24 Maths-Whizz delivery in schools from Sep ’24 – Jul ‘25 Maths-Whizz share periodic platform usage reports with NFER throughout delivery period	Whizz Education
	Statistical Analysis Plan (SAP) First draft of SAP submitted in Sep ’24, edits based on EEF, peer reviewer and delivery partners’ feedback, final SAP published in Jan ‘25	Gemma Schwendel

Dates	Activity	Staff responsible/ leading
	Implementation and Process Evaluation (IPE) Observation of teacher training and onboarding in Aug – Sep ’24 IPE instrument development in Dec ’24 – Jan ’25 IPE activities (case study visits, ESP focus groups) in Jan – Mar ‘25	Katherine Aston
	NPD application Submit NPD application in Oct ‘24	Gemma Schwendel
	Endpoint assessments Endpoint assessments by NFER Test Administrators in Jun – Jul ‘25	Kathryn Hurd, Jishi Jose
Jul – Aug ‘25	**Data cleaning**	Kathryn Hurd
Oct – Dec ‘25	**Analysis** Access matched pupil-level dataset on SRS, complete primary and additional analyses	Gemma Schwendel, Aarti Sahasranaman
Jan – May ‘26	**Reporting** Submit first draft of report in Jan ’26 Three rounds of report reviews from Feb – Apr ’26 Final report published in May ‘26	Aarti Sahasranaman
Aug ‘26	**Data archiving**	Gemma Schwendel

References

Adelson, J.L. and McCoach, D.B. (2011) 'Development and psychometric properties of the math and me survey: measuring third through sixth graders' attitudes toward mathematics', *Measurement and Evaluation in Counseling and Development*, 44(4), pp. 225–247. Available at: https://doi.org/10.1177/0748175611418522.

Borra, C., Lacovou, M. and Sevilla, A. (2021) *Adolescence development and the math gender gap, IZA DP No. 14077*. Available at: https://docs.iza.org/dp14077.pdf (Accessed: 4 April 2024).

Cansiz, M.A. and Tabak, S. (2018) 'Turkish adaptation of math and me survey: a validity and reliability study', *European Journal of Educational Research*, 7(3), pp. 707–714. Available at: https://doi.org/10.12973/eu-jer.7.3.707.

Cilesiz, S. and Greckhamer, T. (2020) 'Qualitative comparative analysis in education research: its current status and future potential', *Review of Research in Education*, 44(1), pp. 332–369. Available at: https://doi.org/10.3102/0091732X20907347.

Department for Education (2022a) *Implementation of education technology in schools and colleges*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1107808/Implementation_of_education_technology_in_schools__and_colleges.pdf (Accessed: 1 October 2023).

Department for Education (2022b) *School-led tutoring guidance*. Available at: https://assets.publishing.service.gov.uk/media/62601734e90e072a014d517d/Updated_School-Led_Tutoring_Guidance_.pdf (Accessed: 5 March 2024).

Department for Education (2023) *Schools, pupils and their characteristics, Academic year 2022/23*, *GOV.UK*. Available at: https://explore-education-statistics.service.gov.uk/find-statistics/school-pupils-and-their-characteristics (Accessed: 4 March 2024).

Education Endowment Foundation (2019) *Putting evidence to work: a school's guide to implementation*. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/eef-guidance-reports/implementation/EEF_Implementation_Guidance_Report_2019.pdf?v=1696494696 (Accessed: 1 October 2023).

Education Endowment Foundation (2022) *Best evidence on impact of COVID-19 on pupil attainment*. Available at: https://educationendowmentfoundation.org.uk/guidance-for-teachers/covid-19-resources/best-evidence-on-impact-of-covid-19-on-pupil-attainment (Accessed: 6 March 2024).

Harrison, W. and Higgins, S. (2023) *A global evidence review for policymakers: findings from meta-analyses or tutoring*. Available at: https://www.edt.org/research-and-insights/a-global-evidence-review-for-policymakers-findings-from-meta-analyses-of-tutoring/#:~:text=02%2F08%2F2023-,A%20global%20evidence%20review%20for%20policymakers,from%20meta%2Danalyses%20of%20tutoring&text=Tutoring%20is%20widely%20acknowledged%20as,of%20the%20Covid%2D19%20pandemic. (Accessed: 6 October 2023).

Hodge, L., Little, A. and Weldon, M. (2021) *GCSE attainment and lifetime earnings*. Available at: https://assets.publishing.service.gov.uk/media/60c36f0cd3bf7f4bd11a2326/GCSE_Attainment_and_Lifetime_Earnings_PDF3A.pdf (Accessed: 6 March 2024).

Mavrikis, M., Schlepps, D.S. and Mubeen, J. (2018) 'Short-term efficacy and usage recommendations for a large-scale implementation of the Math-Whizz Tutor', in C. Penstein Rosé, R. Martínez-Maldonado, H.U. Hoppe, R. Luckin, M. Mavrikis, K. Porayska-Pomsta, B. McLaren, and B. Du Boulay (eds) *Artificial Intelligence in Education (AIED 2018)*. Cham: Springer. Available at: http://link.springer.com/10.1007/978-3-319-93846-2_85 (Accessed: 29 September 2023).

Menzies, L., Ramaiah, B. and Boulton, C. (2021) *A space for maths: exploring the need for maths tutoring and the potential role of Third Space Learning*. Available at: https://cfey.org/wp-content/uploads/2021/09/A-Space-for-Maths-Sept-2021-for-publication-amended.pdf (Accessed: 6 March 2024).

Outhwaite, L., Early, E., Herodotou, C. and Van Herwegen, J. (2022) *Can maths apps add value to young children's learning?* Available at: https://www.nuffieldfoundation.org/wp-content/uploads/2022/05/Can-Maths-Apps-Add-Value-to-Young-Childrens-Learning-A-Systematic-Review-and-Content-Analysis_Web_final_v2.pdf (Accessed: 6 March 2024).

Paz-Albo, J. and Hervás-Escobar, A. (2023) 'Spanish adaptation of the math and me survey in primary education: Measuring second and fourth graders' attitudes toward mathematics', *Cogent Education*, 10(1), p. 2204707. Available at: https://doi.org/10.1080/2331186X.2023.2204707.

Renaissance Learning (2014) *STAR Assessments: the science of STAR*. Available at: https://www.renlearn.co.uk/wp-content/uploads/2014/03/science-of-star.pdf (Accessed: 5 March 2024).

Renaissance Learning (2017) *Relating results from Renaissance Star Reading and Renaissance Star Maths to the Key Stage 2 Standardised Attainment Tests (SATs)*. Available at: https://renaissance.widen.net/view/pdf/f2hkqeg6qy/UK-DMAP-Techinical-Report-12pp-Final.pdf?t.download=true&u=zceria (Accessed: 5 March 2024).

Renaissance Learning and Education Policy Institute (2021) *Understanding progress in the 2020/2021 academic year: interim findings*. Available at: https://assets.publishing.service.gov.uk/media/6239cbb1e90e07799a20bc0e/Understanding_progress_in_the_2020_21_academic_year_Interim_findings_Jan2021.pdf (Accessed: 6 March 2024).

Sewell, J., Sainsbury, M., Pyle, K. and Styles, B. (2007) *Renaissance learning equating study*. Available at: https://www.renlearn.co.uk/wp-content/uploads/2010/12/SM%20-%20Renaissance%20Learning%20Equating%20Study%20Report.pdf (Accessed: 6 March 2024).

Singh, A., Uwimpuhwe, G., Vallis, D., Akhter, N., Coolen-Maturi, T., Higgins, S., Einbeck, J., Culliney, M. and Demack, S. (2023) *Improving power calculations in educational trials*. London: EEF. Available at: https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluation/eef-evaluation-reports-and-research-papers/methodological-research-

and-innovations/improving-power-calculations-in-educational-trials (Accessed: 2 October 2023).

Weidmann, B., Allen, R., Bibby, D., Coe, R., James, L., Plaister, N. and Thomson, D. (2022) *COVID-19 disruptions: Attainment gaps and primary school responses*. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/FFT-COVID-19-disruptions-attainment-gaps-and-primary-school-responses-report-Feb-2022.pdf?v=1709700201 (Accessed: 6 March 2024).

Whizz Education (2021) *Whizz Education proof pack*. Available at: https://www.whizzeducation.com/wp-content/uploads/2022/02/2021-Proof-Pack-Whizz.pdf (Accessed: 6 March 2024).

YouGov (2014) 'National numeracy survey results'. Available at: http://cdn.yougov.com/cumulus_uploads/document/b6j344revv/YouGov-Survey-Results-National-Numeracy-140227.pdf (Accessed: 6 March 2024).

Appendix – R syntax for sample size calculations

```
#=============================================================================
# Script Name: Power_Calculations_EEMW.R
# Script Purpose: Perform power calculations for EEMW
#
# Author: Gemma Schwendel
# Date Created: 19/09/2023
# Notes:
#=============================================================================
# Revision Log
#=============================================================================
# Date              |  Reason for revision
# 02/10/2023        | Updated to use new parameters for ICC and pre-post
#                     as per the EEF paper:
#
https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/metho
dological-research-and-innovations/Work_Package_2023-
WP6_18_09_2023_FINAL.pdf?v=1696233531 =
# =============================================================================

# Set working directory
setwd("C:\\Users\\schwg\\OneDrive - National Foundation for Educational
Research\\Documents\\EEMW")
library(PowerUpR)
library(ggplot2)

rm(list=ls())

################################### Assumptions ############################
### The model used is a 3-level block randomisation with random effects
across level 3 blocks
## Level One: Pupil
## Level Two: Year Group
## Level Three: School

power = 0.8
alpha = 0.05
ICC = 0.109 ## ICC KS2 maths for NPD whole
ICC2 = ICC*(2/3) # Assumed Overall ICC 0.11 with 2/3 accounting for
variation at class level
ICC3 = ICC*(1/3) # Assumed Overall ICC 0.11 with 1/3 accounting for
variation at school level
om = 0.5 # Assumed treatment effect heterogeneity = 0.5
ppn = 0.5 # Assume half go into each treatment arm
g3 = 1 # number of covariates at level 3. Assumed one. Varying this has
little difference
r21 = 0.693^2 # proportion of level 1 variance in the outcome explained by
level 1 covariates
r22 = 0 # proportion of level 2 variance in the outcome explained by level
2 covariates: Assumed no level 2 covariates
r2t3 = 0.497^2 # proportion of treatment effect variance among level 3
units explained by level 3 covariates
n = 26.7 # DfE reported mean for 2022/23
cls = 1.4 # Classes per year group: assumed 1.4: reasonable given ROS
J = 4*cls # Classes per school: assumed 1.4 holds across all year groups
pupil_ret = 0.85
school_ret = 1.0 # Assumed all schools are retained
schools = 100 # Initial starting point
ppn_fsm = 0.246 # Source:
https://lginform.local.gov.uk/reports/lgastandard?mod-metric=2173&mod-
```

```
area=E06000031&mod-group=AllSingleTierAndCountyLaInCountry_England&mod-
type=namedComparisonGroup
MDES = 0.2 # Target MDES
MDES_one = 0.22 # Target MDES for one class
seq_min = 20
seq_max = 150

# Year 2 (KS1) specific parameters
ICC1 = 0.035 ## ICC KS1 maths for NPD whole
ICC21 = ICC1*(2/3) # Assumed Overall ICC 0.036 with 2/3 accounting for
variation at class level
ICC31 = ICC1*(1/3) # Assumed Overall ICC 0.036 with 1/3 accounting for
variation at school level
r211 = 0.350^2 # For KS1 maths from NPD whole
r2t31 = 0.365^2 # For KS1 maths from NPD whole

##########################################################################
########################################
## All pupils, 100 schools
mdes.bcra3r2(power=power, alpha=alpha, two.tailed=TRUE,rho2=ICC2,
rho3=ICC3, omega3=om,
             p=ppn, g3=g3, r21=r21, r22=r22, r2t3=r2t3, n=n*pupil_ret, J=J,
K=schools*school_ret)

## FSM pupils, 100 schools
mdes.bcra3r2(power=power, alpha=alpha, two.tailed=TRUE,rho2=ICC2,
rho3=ICC3, omega3=om,
             p=ppn, g3=g3, r21=r21, r22=r22, r2t3=r2t3,
n=n*ppn_fsm*pupil_ret, J=J, K=schools*school_ret)

## One class, 100 schools: effectively, a KS2 class
mdes.bcra3r2(power=power, alpha=alpha, two.tailed=TRUE,rho2=ICC2,
rho3=ICC3, omega3=om,
             p=ppn, g3=g3, r21=r21, r22=r22, r2t3=r2t3, n=n*pupil_ret, J=1,
K=schools*school_ret)

## One class (KS1), 100 schools, using KS1 parameters as defined by new EEF
paper
mdes.bcra3r2(power=power, alpha=alpha, two.tailed=TRUE,rho2=ICC21,
rho3=ICC31, omega3=om,
             p=ppn, g3=g3, r21=r211, r22=r22, r2t3=r2t31, n=n*pupil_ret,
J=1, K=schools*school_ret)

## One year group, 100 schools: effectively, these are KS2 classes
mdes.bcra3r2(power=power, alpha=alpha, two.tailed=TRUE,rho2=ICC2,
rho3=ICC3, omega3=om,
             p=ppn, g3=g3, r21=r21, r22=r22, r2t3=r2t3, n=n*pupil_ret,
J=cls, K=schools*school_ret)

## Year Two (KS1), 100 schools, using KS1 parameters as defined by new EEF
paper
mdes.bcra3r2(power=power, alpha=alpha, two.tailed=TRUE,rho2=ICC21,
rho3=ICC31, omega3=om,
             p=ppn, g3=g3, r21=r211, r22=r22, r2t3=r2t31, n=n*pupil_ret,
J=cls, K=schools*school_ret)

### Now generate graphs & output dataset
grDat<-data.frame(Scenario=factor(rep(rep(c("All Years","FSM Pupils","One
KS2 Class","One KS1 class","KS2 year","KS1 year"),
                  each=length(seq(seq_min,seq_max,1))),1),levels=c("All
Years","FSM Pupils","One KS2 Class","One KS1 class","KS2 year","KS1
year")),
```

```
                nSetting=rep(seq(seq_min,seq_max,1),6),


MDES=c(sapply(seq(seq_min,seq_max,1),function(i){mdes.bcra3r2(power=power,
alpha=alpha, two.tailed=TRUE,rho2=ICC2, rho3=ICC3, omega3=om,

p=ppn, g3=g3, r21=r21, r22=r22, r2t3=r2t3, n=n*pupil_ret, J=J,
K=i*school_ret)$mdes[1,1]}),

sapply(seq(seq_min,seq_max,1),function(i){mdes.bcra3r2(power=power,
alpha=alpha, two.tailed=TRUE,rho2=ICC2, rho3=ICC3, omega3=om,

p=ppn, g3=g3, r21=r21, r22=r22, r2t3=r2t3, n=n*ppn_fsm*pupil_ret, J=J,
K=i*school_ret)$mdes[1,1]}),

sapply(seq(seq_min,seq_max,1),function(i){mdes.bcra3r2(power=power,
alpha=alpha, two.tailed=TRUE,rho2=ICC2, rho3=ICC3, omega3=om,

p=ppn, g3=g3, r21=r21, r22=r22, r2t3=r2t3, n=n*pupil_ret, J=1,
K=i*school_ret)$mdes[1,1]}),

sapply(seq(seq_min,seq_max,1),function(i){mdes.bcra3r2(power=power,
alpha=alpha, two.tailed=TRUE,rho2=ICC21, rho3=ICC31, omega3=om,

p=ppn, g3=g3, r21=r211, r22=r22, r2t3=r2t31, n=n*pupil_ret, J=1,
K=i*school_ret)$mdes[1,1]}),

sapply(seq(seq_min,seq_max,1),function(i){mdes.bcra3r2(power=power,
alpha=alpha, two.tailed=TRUE,rho2=ICC2, rho3=ICC3, omega3=om,

p=ppn, g3=g3, r21=r21, r22=r22, r2t3=r2t3, n=n*pupil_ret, J=cls,
K=i*school_ret)$mdes[1,1]}),

sapply(seq(seq_min,seq_max,1),function(i){mdes.bcra3r2(power=power,
alpha=alpha, two.tailed=TRUE,rho2=ICC21, rho3=ICC31, omega3=om,

p=ppn, g3=g3, r21=r211, r22=r22, r2t3=r2t31, n=n*pupil_ret, J=cls,
K=i*school_ret)$mdes[1,1]})))

tiff("EEMW_SampleSize_revisedEEF_NPDsamp.tiff",width=8,height=5,unit="in",r
es=400,compression="lzw")
print(ggplot(grDat, aes(x=nSetting,y=MDES,group=Scenario,colour=Scenario,))
+
        geom_line()+
        geom_hline(yintercept = MDES,linetype="dashed")+
        geom_hline(yintercept = MDES_one,linetype="dashed",col="red")+
        labs(x="Number of settings (total)",colour=NULL)+
        theme_bw()
)
graphics.off()
write.csv(grDat,"EEMW_SampleSize_revisedEEF_NPDsamp.csv",row.names=F,na="")
```