# *Improving Working Memory plus Arithmetic*
# Evaluation Protocol

## Evaluator (institution): RAND Europe
## Principal investigator: Elena Rosa Brown

Template last updated: August 2019

| | |
|---|---|
| PROJECT TITLE | Improving Working Memory plus Arithmetic |
| DEVELOPER (INSTITUTION) | University of Oxford |
| EVALUATOR (INSTITUTION) | RAND Europe |
| PRINCIPAL INVESTIGATOR(S) | Elena Rosa Brown |
| PROTOCOL AUTHOR(S) | Elena Rosa Brown, Lillian Flemons, Emma Leenders, Simon Larmour, Annemari de Silva |
| TRIAL DESIGN | Two-arm cluster randomised controlled trial with random allocation at school level |
| TRIAL TYPE | Effectiveness |
| PUPIL AGE RANGE AND KEY STAGE | 6-8, KS2 (Year 3) |
| NUMBER OF SCHOOLS | 200 - 240 |
| NUMBER OF PUPILS | 2,400 |
| PRIMARY OUTCOME MEASURE AND SOURCE | Number Skills (BAS3 number skills test) |
| SECONDARY OUTCOME MEASURE AND SOURCE | Working memory (Working Memory battery, central executive sub-texts)<br>Wider maths attainment (GL Assessment's Progress Test in Mathematics 8)<br>Attention & behaviour (SNAP-IV Teacher Attention Rating Scale) |

## Protocol version history

| VERSION | DATE | REASON FOR REVISION |
|---|---|---|
| 1.0 [*original*] | | N/A |

## Table of contents

## Study rationale and background

Working memory (WM) is the ability to keep information in mind and to use it to guide behaviour without the support of external cues. There are different theoretical models of WM, but the two models that underpin most current research (Baddeley, 2010; Ricker, AuBuchon & Cowan, 2010) have two features in common: a central executive, which controls attention, and a long-term memory component, which in Baddeley's model is termed "episodic buffer". The central executive is an attention control system, responsible for the "selective description of information that is relevant to current thought or behaviour" (Duncan, 2006). The long-term memory component influences the selection and maintenance of the information in an available state to support conscious activity (Ricker et al., 2010); it influences, for example, how information is organised during processing. Duncan (2006) argued that, from the neuropsychological perspective, WM and attention are intrinsically related functions.

Working memory underpins cognitive development and is a fundamental element of learning (Cowan, 2014; Baddeley, 2012) and has been shown to be linked to performance on maths-related tasks, including basic arithmetic and multiplication (Baddeley 2012; Seitz & Schumann-Hengsteler, 2000). For example, in arithmetic, we often decompose numbers and carry out several steps over the course of the calculation, keeping track of the intermediary results; this decomposition involves 'chunking' the numerical information in an effective way (i.e., thinking of 249 as 200+40+9). In order to understand some basic number concepts, we need to keep at least four pieces of information in WM. For example, to appreciate the inverse relation between addition and subtraction represented in the expression 15+23-23=15). A previous efficacy trial funded by the EEF (Wright et al., 2019) found that approximately 25% of the pupils could only hold two items in their WM, while for many mathematical tasks, students need to retain at least three items.

The Improving Working Memory + Arithmetic (IWM+A) intervention is for six- to eight-year-old children who show low attainment in arithmetic at the end of Key Stage 1 (KS1) (children aged five to seven). It combines a WM, central executive intervention with an arithmetic intervention, and was designed to improve WM span and the way numbers and arithmetic operations are organised in long-term memory. Thus the intervention targets two components of WM, the central executive and the episodic buffer, which involves an input from long-term memory. The activities were designed using research findings about how attention can be promoted and how numerical information can be stored in long term memory and processed in effective ways. During the WM intervention, children learn strategies for keeping information in mind so that their increased WM capacity allows them to learn basic number concepts, such as additive composition and the inverse relation between operations. As their WM improves, children can learn these concepts and represent numbers and operations in their long-term memory in a new way. The combination of WM with an additional component that aims to enhance the understanding of underlying principles of arithmetic is particularly promising, as Craick & Lockhart famously showed that deeper, more elaborate processing leads to better learning (Craik & Lockhart, 1972). This is supported by previous studies by the Oxford University Team (Nunes et al., 2008; Nunes et al., 2014). In addition, there is evidence of the arithmetic element of the intervention being successful, with previous evaluations of similar approaches from the same Oxford University Team showing positive impact on pupils' numeracy ability (Worth et al., 2015).

A previous EEF efficacy trial was conducted of Improving Working Memory Plus (IWM+) (the same intervention as IWM+A, but with a different name) in which TAs who delivered the intervention were trained by Oxford University. The study found an impact of additional three months progress in number skills compared to children in the control condition (d=0.24) (Wright et al., 2019). Based on this evidence, Oxford University was granted funding by the EEF to deliver IWM+A to more schools, using a train-the-trainer model, with Oxford University training Teacher Leaders (TL) who in turn will train Teaching Assistants (TAs) across 12 regions in England.

It should be noted that findings for pupils eligible for free school meals (FSM) were more mixed in the IWM+ trial, suggesting that FSM students made gains in WM but not in maths (d=-0.02). Other studies of the effects of working memory and cognitive training programmes on pupils' attainment suggests mixed findings are quite common (Shipstead et al., 2010), while another study shows WM training yields short-term improvements (Melby-Lervåg & Hulme, 2013).

Given the mixed evidence on the effects of the programme on the number skills of FSM pupils and the move to a 'train-the-trainer' approach, it is crucial that the next evaluation of IWM+A looks closely at impact on pupils, particularly for FSM pupils. The proposed effectiveness trial will build on the IWM+ trial by ensuring the trial is powered to detect an effect on FSM pupils, as well as looking at how the intervention will be delivered at scale.

## Intervention

**Name:** Improving Working memory plus Arithmetic (IWM+A)

**Why (theory /rationale):**

IWM+A is a multi-factorial regime intervention, targeting both WM and arithmetic skills (Nunes et al., 2021). WM has been found to be a robust longitudinal predictor of mathematic achievement in primary school, even after stringent controls for such factors as age and general cognitive skill (Friso-Van Den Bos et al., 2013; Peng, Namkung, Barnes & Sun, 2016; Liew et al., 2010). The strength of the relation between WM and mathematics achievement is moderated by the measure of WM used, with the central executive tending to show a stronger association (De Smedt et al., 2009; Nunes et al., 2007). It is also moderated by type of mathematical task, for example, word-problem solving and whole-number calculations showed the strongest relation with WM whereas geometry showed the weakest relation with WM (Peng et al., 2016). Correlation is stronger amongst children with mathematical difficulty than amongst typically developing children (Peng et al., 2016).

It has also been shown that WM can be improved with training (von Bastian & Oberauer, 2014), although this improvement may not generalise to other skills essential for academic learning, such as arithmetic, (Melby-Lervåg & Hulme, 2013). The lack of transfer of gains in WM task to arithmetic tasks in a WM intervention may be a consequence of the connection between WM and long-term memory (Baddeley, 2000; Cowan, 1988; Ricker et al., 2010). An intervention that combines WM and arithmetic can lead to changes in long term memory and have a positive impact children's arithmetic learning.

The IWM+A intervention was designed for children aged six to eight years who show low attainment in maths at the end of KS1, and thus are most likely to benefit from a WM intervention. Its aims are to improve the executive component of WM and to promote children's skills in the organisation of information about numbers and about the operations of addition and subtraction in long-term memory.

TAs support children's improvement of the attention control system achieved by scaffolding children's use of strategies (i.e. verbal rehearsal, visuo-spatial organisation of information). TAs scaffold through modelling and practising alongside the children, when required, but relinquish control when appropriate. TAs further support children by, asking them to verbalise the strategies, giving positive feedback on the use of the strategies and attributing success to the use of the strategies. The improvement in WM span, in turn, supports the children to make different connections between numbers and arithmetic operations, with the aim of promoting reorganisations in long-term memory that will facilitate further learning of number skills. For example, WM plays a role in learning about additive composition because the child must be able to recall three numbers while operating on them (Nunes et al. n.d.). Previous research has shown that an intervention that aims to promote conceptual understanding of numerical information by children with difficulties is more effective when it is supplemented by a WM component than when it is delivered on its own (Barahmand, 2008).

The games involved in the intervention were also designed to involve complex sequences of activities and the presentation of distractors in order to promote improvement in selective attention (Duncan, 2013). Attention and working memory are intrinsically related functions (Duncan, 2006), as attention involves the 'selective description of information that is relevant to current thought or behaviour' (Duncan 2006), while working memory is the maintenance of this information in an available state to support conscious activity (Ricker et al. 2010). This means that effective attention is a necessary precursor to effective working memory.

**What**: The intervention is delivered by trained teachers or TAs, who are expected to scaffold the use of strategies in the WM and number tasks. Both components, the WM and the Arithmetic intervention, include two types of activity:

- TA led activities, during which scaffolding of the strategies takes place, and
- web-based games, which the child plays independently.

The content of the activities varies. In the WM games, the children need to recall different words, digits, animals, colours, or letters. The demands on WM also vary across tasks. A number of tasks are provided, including tasks where children must:

- process the information in a sentence and recall the last word;
- reverse the order in which the items were presented;
- inhibit the recall of rehearsed items because they are no longer required for the task.

The WM tasks are presented at levels that are progressively more challenging for each child. The numerical tasks require the use of counting on, additive composition, and the inverse relation between operations in different context and types of game. The level of difficulty is also increased progressively.

**Materials:** There are a number of materials to support delivery of IWM+A**.**

Games:

- WM games:
    - There are three TA-led games (words, colours, and missing digits). In order to deliver the TA-led games, the TA uses PowerPoint slides designed for the activities. The TA is asked to record the children's accuracy and to proceed with the games according to criteria for success. The TA is asked to play at least two different games in each session.
    
      There are three computer-adaptive web-based games for children to play independently (animals, letters, and digit sequences). The children record the web-based games that they have played on a record sheet. Success in the web-based WM games leads to access to a different type of game (bonus games); the children also receive certificates, available for the TA to download and print, which the children can be awarded at an event in school (e.g. assembly) and take home.
- Arithmetic games:
    - There are also TA led and web-based games. In order to deliver the TA-led games, the TA uses PowerPoint slides designed for the activities; some games also require the use of coins and cards designed for the activities. During the TA-led games, the TA is expected to record the child's answers and strategies in order to promote their use consistently. For some of the TA led activities, the child also has an answer sheet. The child has a record sheet for the web-based games. The games are self-paced as they are not computer adaptive

Two implementation handbooks are provided, one for the WM component and one for the arithmetic component of the programme. These handbooks supplement the training by providing detailed instructions for the games, and guidance for optimal delivery of the intervention. Following Phase 1, feedback on the Handbook from the Teacher Leaders and observers will be collated. This feedback will subsequently be used to optimise the handbooks for use with TAs in Phase 2. There will also be Summary Sheets (one for each game) that can act as a quick reference to the TA during the intervention.

**Procedures:** IWM+A is delivered in one-to-one sessions, offered in addition to dedicated numeracy lesson time, with TAs working with two pupils over the course of an hour. TAs work directly with one child for half an hour while a second child plays targeted computer games in the same room. After half an hour the children switch activities. During the one-to-one element, TAs demonstrate and practice strategies with children, which are also reinforced through adaptive online games. The children access the targeted games online with an individualised login. The WM programme is individually paced, so that each child is automatically moved to the next level of the game when appropriate, in order to keep the level of challenge suitable for each child. The arithmetic games are also self-paced but are not computer-adaptive. Children receive a record sheet that is used to record the games they played; TAs guide them when they log in to the arithmetic games.

Children progress at their own pace through the sessions, aided by TAs, with the one-to-one nature of delivery ensuring the intervention is personalised to individual children. TAs are able to access extra materials to support further learning. The minimum requirements to complete the programme is five WM combined with four arithmetic sessions; though five WM combined with five arithmetic sessions is considered optimal.

**Who provided:** To enable the intervention to reach more schools, the Oxford University team will be training 12 TLs to deliver the intervention across 12 regions in England (see Participant Selection for more details). TLs recruited by Oxford are former Every Child Counts trainers and were selected because they (a) have a wide range of experience of training teachers and TAs in mathematics interventions; (b) have a thorough understanding of the primary mathematics curriculum; (c) understand how children learn mathematics and how to support struggling learners; (d) Know how to build a learning community with a shared vision and values; and (e) have successful experience of supporting schools to implement change. Oxford University will train TLs in Phase 1 (October 2020 – July 2021). The trainings are intended to be in-person, barring COVID-19 restrictions, which necessitated that the first two training sessions took place remotely. TLs will practise delivery in a small number of schools (not included as part of the Phase 2 trial) with three pairs of children per school: one pair in the morning, one pair after the morning break, and one pair in the afternoon. Oxford University will be closely involved undertake observations to monitor implementation and implement professional development conversations with each TL as part of their training in Phase 1. TLs will receive two additional training sessions in Phase 2 to support their ability to deliver training, to monitor quality of the TA delivery, and to provide professional development to TAs. The two training sessions will be convened immediately before TLs deliver training to TAs and Link Teachers - a teacher in the school responsible for supporting the TA.

In Phase 2 of the evaluation (September 2021 – June 2022), TLs will train one Link Teacher and one TA nominated by each participating school. Schools are asked to nominate a TA who likes teaching mathematics, has experience with teaching children who find mathematics challenging, and has sufficient familiarity with computers to use the PowerPoint materials and guide the children in the use of the web-based games.

TAs are trained alongside Link Teachers over two days of training and receive two implementation handbooks, one for each part of the intervention programme. The first day focuses on the WM aspect and the second training day focuses on arithmetic. Each TA also receives a visit from the TL responsible for the school to support delivery and provide quality assurance. TAs are trained in how to use IWM+A materials (e.g. the TA delivered games and the online games) and how to teach strategies that have been shown to improve children's WM span and arithmetic skills. TAs learn how to implement guided rehearsal to teach WM strategies to children and how to praise the use of rehearsal rather than success in the tasks. Oxford University will provide the IWM+A materials, which include a handbook to be used during both training and implementation, and access to the website for games and supporting materials.

**How:** All schools will only be able to nominate one TA for training, who will be expected to deliver to ten children. TAs work directly with one child for half an hour, implementing the TA led games, while a second child plays targeted computer games in the same room. After half an hour the children switch activities.

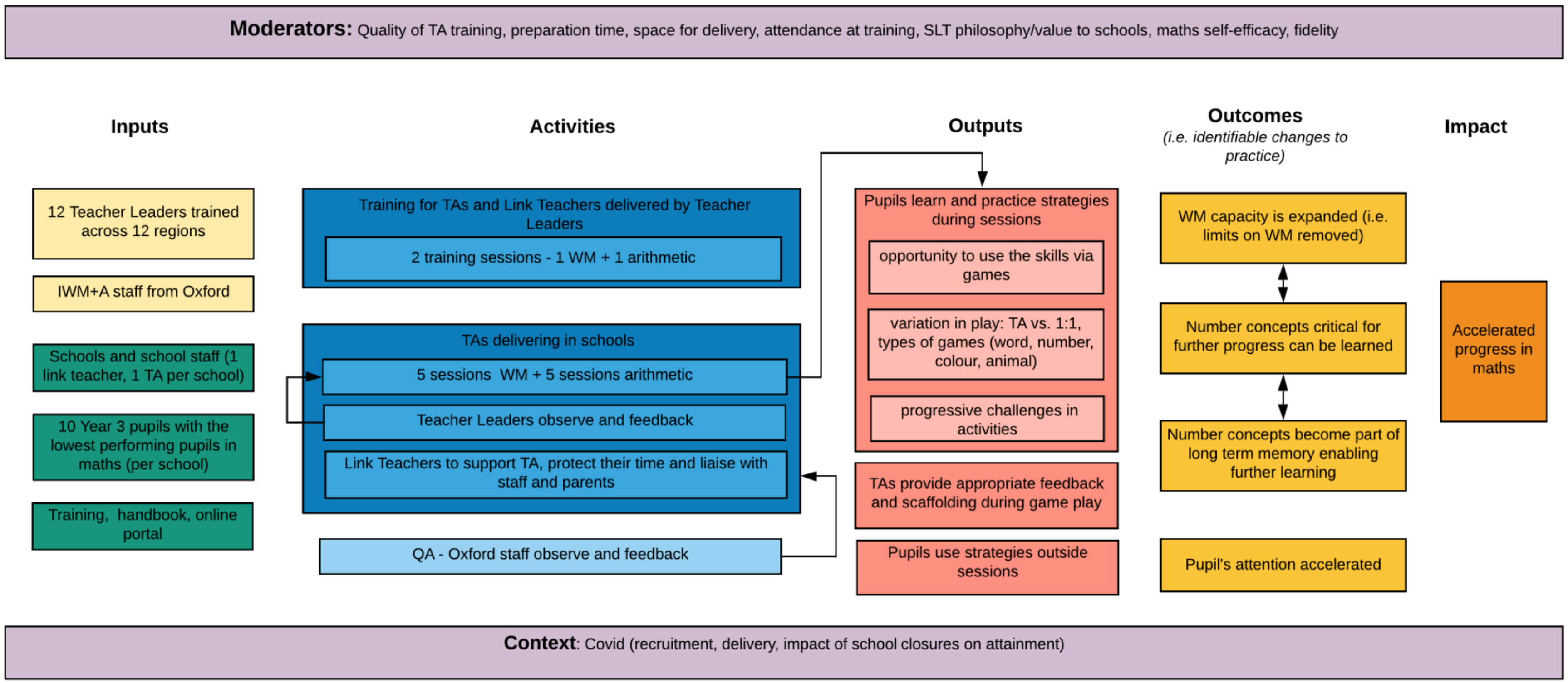
**Where (locations):** IWM+A is designed to be delivered in schools in areas available for small group or one to one work, such as empty classrooms. It is recommended that the school provide a quiet, suitable space free from distractions.

**When and how much (dosage):** The intervention is delivered for one hour every week over the course of ten weeks. The first five weeks of the intervention focus on improving WM and the final five weeks focus on number and arithmetic operations, developing the pupil's ability to understand relations between numbers (e.g. additive composition) and between arithmetic operations (e.g. understanding the inverse relationship between addition and subtraction). Children only move on to the arithmetic component after participating in the WM intervention.

**Tailoring (adaptation):** TLs were trained remotely due to social distancing restrictions imposed in January of 2021. The WM and Arithmetic training sessions were provided online as well as through the individual online sessions during which professional development conversations between the Oxford team and the TLs take place. There are no changes planned to the content or delivery model of the TA training sessions or individual professional development as a result of the COVID-19 pandemic.

The Theory of Change  (Figure 1) presents an overview of IWM+A.

***Figure 1 Theory of Change of IWM+A***



## Impact evaluation

### *Research questions*

The primary research question of this project is:

1. What is the difference in number skills measured by the number skills subtest of the British Ability Scales (BAS3) of pupils in schools receiving IWM+A in comparison to those pupils in control schools receiving business as usual?

The secondary research questions of this project are:

2. What is the difference in maths attainment measured by Progress Test in Maths (8) of pupils in schools receiving IWM+A in comparison to those pupils in control schools receiving business as usual?
3. What is the difference in WM measured by the Working Memory Battery Test (Central Executive Component) of pupils in schools receiving IWM+A in comparison to those pupils in control schools receiving business as usual?
4. What is the difference in attention and behaviour measured by an adapted 15-item SNAP-IV Teacher Attention Rating Scale of pupils in schools receiving IWM+A in comparison to those pupils in control schools receiving business as usual?
5. What is the difference in number skills measured by the number skills subtest of the British Ability Scales (BAS3) of FSM pupils in schools receiving IWM+A in comparison to FSM pupils in control schools receiving business as usual?

### *Design*

**Table 1: Trial design**

| | | |
|---|---|---|
| **Trial design, including number of arms** | | Two-arm cluster randomised controlled trial with random allocation at school level |
| **Unit of randomisation** | | School |
| **Stratification variables** (if applicable) | | Region: Cambridgeshire, Derbyshire, Devon, East Sussex, Essex, Greater Manchester, Merseyside, Nottinghamshire, South Yorkshire, Suffolk, West Sussex, West Yorkshire |
| **Primary outcome** | variable | Number skills |
| | measure (instrument, scale, source) | British Ability Scales, Third edition (Number skills test, GL Assessments) |
| **Secondary outcome(s)** | variable(s) | • Working memory<br>• Maths attainment<br>• Attention and behaviour |
| | measure(s) (instrument, scale, source) | • Working Memory Battery (WMB) (Listening Recall, Counting Recall, Backward Digit Recall subtests, Pickering & Gathercole, 2001)<br>• Progress Test in Maths (PtM) (Level 8, GL Assessments) |

| | | |
|---|---|---|
| | | • SNAP-IV Teacher Attention Rating Scale (adapted 15-item instrument [1] (Swanson et al., 2001; Bussing et al., 2008, Hall et al., 2020) |
| **Baseline for primary outcome** | **variable** | Maths attainment |
| | measure (instrument, scale, source) | Early Years Foundation Stage Profile (EYFSP) (Numbers, and Shape, Space and Measures variables, NPD) |
| **Baseline for secondary outcome** | **variable** | Attention and behaviour |
| | measure (instrument, scale, source) | SNAP-IV Teacher Attention Rating Scale (adapted 15-item instrument, Swanson et al., 2001) |

This evaluation is designed as a region-stratified, two-arm, cluster randomised, effectiveness trial as outlined in Table 1. Random allocation will be at school level to avoid contamination, with schools randomly allocated to the treatment condition delivering IWM+A and schools randomly allocated to control condition delivering business as usual.

Randomisation will be stratified within the different geographical areas to balance study arms across delivery locations (which are geographically based). This is an effectiveness trial, building on a previous EEF efficacy trial (Wright et al. 2019). Outcomes mirror those used in the efficacy trial, including number skills (primary outcome, measured using the BAS3), WM (secondary outcome, measured using WMB) and attention and behaviour (secondary outcome, measured using SNAP-IV). Maths attainment will also be measured in this trial to understand how IWM+A affects maths attainment against the national curriculum (secondary outcome, measured using PtM8). More details on outcome measures are discussed in the 'Outcome measures' section below.

The evaluation will be comprised of two phases: Phase 1 focuses on the train-the-trainer aspect; and Phase 2 is the main trial.

### *Randomisation*

Randomisation will be stratified by region with schools as the unit of randomisation, and (Year 3) pupils as the unit of analysis. Given delivery and recruitment are organised regionally (i.e. with one TL responsible for recruiting and delivering in a single region) we will stratify by region. Randomising at the school level also reduces the chances of contamination between randomised groups as TAs will be delivering to all identified children in the school.

Stratifying at the region/trainer level is particularly relevant given the regional nature of delivery with one TL delivering the training in each region (see the 'Intervention' section for more details). Regions to be included are: Cambridgeshire, Derbyshire, Devon, East Sussex, Essex, Greater Manchester, Merseyside, Nottinghamshire, South Yorkshire, Suffolk, West Sussex, West Yorkshire. Stratifying by region will ensure that TLs in each region have a number of schools they will be able to train. We will use a package in stata (randtreat) to avoid unequal allocation due to an uneven number of schools within each region.

Randomisation will occur with a 50:50 allocation to treatment and control. Schools allocated to treatment will receive IWM+A training and be expected to deliver the IWM+A programme, while schools allocated to control will be expected to carry on with business as usual (BAU). Schools assigned to the control group will be given £1000 on completion of all endline assessments. These funds are to be used at the discretion of the school and could be used to buy an intervention programme of their choice, including IWM+A from September 2022. The value of the incentive is the same that was used in the efficacy trial.

[1] We will use the same adapted scale that was used in the efficacy trial.

As IWM+A is a targeted intervention, schools will only be eligible for randomisation once target children have been identified (see Pupil Eligibility Criteria) and all requirements for randomisation have been completed (e.g. providing student identifiers and teacher scores on SNAP-IV, confirming information sheets have been shared with parents and carers) and data checked for completeness by the Oxford University team.

Randomisation will be conducted in Stata by a member of the evaluation team, who will be blind to treatment allocation. The code used to randomise schools as well as all relevant variables will be saved and made available if requested. The trial allocation will be recorded in Excel and communicated to the implementation team in a PDF file to prevent editing.

## *Participants*

Schools will be recruited on a regional basis by TLs with support from Oxford University. TAs, Link Teachers, and pupils will be selected by schools with advice from Oxford University to guide selection. Further details on these are provided below.

### *Schools*

Schools must be located in one of the regions served by the TLs: Cambridgeshire, Derbyshire, Devon, East Sussex, Essex, Greater Manchester, Merseyside, Nottinghamshire, South Yorkshire, Suffolk, West Sussex, West Yorkshire. As a minimum, schools are required to have at least one class of Year 3 pupils (e.g. one form entry). Schools must be at least single form entry with a class of 30 Year 3 pupils, and, ideally, have a higher than average number of FSM pupils.

In addition, the following exclusion criteria will be observed:

- Schools currently delivering WM or WM+. However, schools that have previously delivered one of these interventions are still eligible.
- Schools currently involved in any other EEF early Key Stage 2 (KS2) numeracy trial in the study period.

Each TL will aim to recruit up to 20 schools in their region (see Sample size calculations for more details).

### *Teaching assistants and Link Teachers*

Schools will be responsible for selecting TAs to deliver the intervention and a link teacher to support the TA. Schools will only be able to nominate one TA regardless of the size of the school. Oxford University will make the Senior Leadership Team in schools aware of the responsibilities of Link Teachers and TAs. Link teachers implement and manage the intervention programme. Their responsibilities are to:

- Organise the collection and submission of consent forms and data;
- Plan an agreed timetable for the programme with the TA and Year 3 class teacher;
- Arrange and secure resources required, including teaching space and access to two computers;
- Act as the TA mentor, supporting the delivery of the programme and ensuring sufficient preparation time;
- Oversee any communication with Year 3 class teacher and parents.

Teaching Assistants deliver the intervention programme. In the TA led activity strand, their responsibilities are to:

- Ensure they have one computer or laptop for delivering the TA-led activities;
- Ensure that all materials are available for the session (TA record sheets, child record sheets, materials for the arithmetic TA led sessions);

- Deliver regular 30-minute sessions each week per child for 10 weeks and supervise at the same time a second child playing the web-based games for 30 minutes;
- Complete record sheets for each child during the delivery of the 3 TA-led activities;
- Teach rehearsal strategies and encourage their use through the TA-led activities;
- Implement all 5 sessions of WM before starting the arithmetic sessions;
- Provide a quiet working area for the child;
- Follow the Handbook.

In the Computer games strand, their responsibilities are to:

- Support the child to access the computer games online;
- Supervise the independent playing of the web-based games;
- Check that each child is using the record sheet and keeping a tick list of games played and completed;
- Oversee children's progress.

***Pupils***

Classroom teachers will be responsible for selecting children to participate in IWM+A before the start of the trial (i.e. pre-randomisation). Pupils will be the ten lowest-attaining students in mathematics at the end of their KS1, according to the teachers' judgements. In the previous trial, teachers were encouraged to consider KS1 results in their nominations.

For the purposes of the trial, children will be identified in September of Year 3 (i.e. at the start of the school year). However, given that Year 3 teachers may not be familiar with their new class, the children's former Year 2 (i.e. end of KS1) teachers will be asked to collaborate with Year 3 teachers to identify appropriate pupils. Teachers will be encouraged to use their own judgement, and may base this on their own evaluation of how children perform in the mathematics lessons and activities. In the efficacy trial teachers were asked to use KS1 assessments, but these were cancelled in 2021 as a result of the impact of COVID-19.

For schools with more than one form entry, teachers of all the classes will cooperate and identify the 10 pupils showing lowest performance in numeracy across the year group. Schools will only be able to nominate 10 children, regardless of the size of the school.

In addition, the following exclusion criteria will be observed:

- Deafness, blindness, and/or physical restrictions that might interfere with a child's ability to use the online games.
- Behavioural problems that might interfere with a child's ability to work independently and in the same room as another child.
- Children whose level of fluency in English would prevent them from engaging with the computer games.

## *Sample size calculations*

**Table 2: Sample size calculations**

| | | TOP RANGE | | LOWER RANGE | |
|---|---|---|---|---|---|
| | | OVERALL | FSM | OVERALL | FSM |
| **Minimum Detectable Effect Size (MDES)** | | 0.172 | 0.221 | 0.188 | 0.242 |
| **Pre-test/ post-test correlations[2]** | level 1 (pupil) | 0.50 | | | |
| | level 2 (school) | N/A | | | |
| **Intracluster correlations (ICCs)** | Level 1 (pupil) | 0.16 | | | |
| | level 2 (class) | N/A | | | |

| | TOP RANGE | | LOWER RANGE | | |
|---|---|---|---|---|---|
| **Alpha** | | 0.05 | | | |
| **Power** | | 0.8 | | | |
| **One-sided or two-sided?** | | Two | | | |
| **Average cluster size** | | 10 | 3 | 10 | 3 |
| **Number of schools** | Intervention | 120 | | 100 | |
| | Control | 120 | | 100 | |
| | **Total** | 240 | | 200 | |
| **Number of pupils** | Intervention | 1,200 | 360 | 1,000 | 300 |
| | Control | 1,200 | 360 | 1,000 | 300 |
| | **Total** | 2,400 | 720 | 2,000 | 600 |

Given the regional nature of delivery (and the potential for TL effects within each region) minimum detectable effect size (MDES) for this study has been calculated using a two-level random assignment design with 240, and 10 pupils per school at Level 1. MDES is calculated at Level 2 as the programme is delivered by teaching assistants. ICC and pre-post-test correlations have been taken from the efficacy trial, using 0.16 and 0.5 respectively (Wright et al. 2019)[2]. Using the PowerUp tool for main effects (Dong & Maynard 2013) and assuming an alpha of 5% and an intended 80% power to detect effects, the trial is powered to detect an effect of 0.172

MDES was also calculated for FSM pupils, using the same assumptions as the overall MDES calculations listed above with an average of three FSM pupils per school. This is based on data from the efficacy trial, which suggests that an average of 36.8% of pupils in the trial were eligible for FSM (using the Ever 6 variable). With these assumptions the trial is powered to detect an effect of 0.221 on FSM pupils.

Given the EEF's focus on FSM children, and the potential for this trial to be able to shed light on approaches that might help narrow the attainment gap, the aim of the trial will be to recruit 240 schools. However, it was acknowledged at the set-up stage that it may be difficult for some trainers to recruit 20 schools in their region. The evaluation team, delivery team, and EEF therefore agreed that the trial would proceed if a minimum of 200 schools had been recruited, although the aim was to recruit 240 schools.

## *Outcome measures*

### Baseline measures

We will be using the Early Years Foundation Stage Profile (EYFSP) numeracy components collected from the National Pupil Database (NPD) as a baseline measure of attainment; a baseline for attention will be measured using the adapted SNAP-IV Teacher Attention Rating Scale (Swanson et al., 2001; Bussing et al., 2008, Hall et al., 2020) completed by teachers prior to randomisation.

*EYFSP*

The EYFSP is a nationally administered measure of ability completed by practitioners (i.e. teachers or other qualified school staff) at the end of Reception when children are between the ages of four and five. Practitioners are asked to make judgements against descriptors and decide whether children 'meet', 'exceed', or are 'emerging'. These judgements are then recorded in the NPD as a value of 1, 2, or 3. There is also the option to mark children as 'not assessed'. The EYFSP is divided across the 17

[2] It should be noted that the efficacy trial used KS1 arithmetic scores as baseline, as opposed to EYFSP suggested here. This was used as the assumption for the correlation in the power calculations as there is no currently published data on EYFSP and BAS3 correlations.

early learning goals (ELGs) of the Early Years Foundation Stage (EYFS), but for the purposes of this trial only the two ELGs related to Mathematics will be used for each pupil, namely: Numbers (MAT_G11), and Shape, space, and measures (MAT_G12). While the underlying mathematical concepts of IMW+A focus on arithmetic, given the limited nature of the data points in the baseline (i.e. three values for each ELG), it was decided to combine both aspects of Mathematics to create a Maths total score (i.e. range of 0 – 6) to provide a measure of increased sensitivity.

For the purposes of this trial the EYFSP completed in the 2018/2019 academic year, when the Year 3 children were in Reception, will be used and data from the measure will be collected via the NPD. Prior to randomisation all schools will be asked to share relevant pupil-level data that will allow researchers to link to data held in the NPD. Only data on pupils in the trial will be requested.

EYFSP was selected as an alternative to administering tests in schools prior to randomisation to reduce the burden on schools. KS1 results would have been preferable given this would parallel the approach in the efficacy trial and the fact that they would be collected in the summer before implementation, however, in light of constraints resulting from the COVID-19 pandemic the 2021 KS1 SATs were cancelled by the Department for Education.

We acknowledge that in using the EYFSP as baseline data, we may not be sufficiently capturing current levels of mathematical attainment, particularly given the disruptions caused by school closures as a result of COVID-19. However, to limit data collection burden on schools it was decided to use EYFSP. If the EYFSP is not correlated with outcome measures (i.e. pre-post-test correlation is equal to 0), the study will still be powered to detect an effect of 0.179 on all pupils, and 0.241 on FSM pupils.

*SNAP-IV Teacher Attention Rating Scale*

A second baseline measure will be collected using the adapted SNAP-IV Teacher Attention Rating Scale (Swanson et al. 2001) completed by teachers in September 2021 prior to randomisation.

The SNAP-IV Teacher Attention Rating Scale is a widely used rating scale for hyperactivity, impulsiveness, and inattention (Swanson, 1992). The scale is organised as a four-point likert scale ranging from 'not at all' to 'very much'. The psychometric properties of an abbreviated version of SNAP-IV were analysed in one paper (Bussing et al., 2008) resulting in the identification of two relevant factors, inattention and hyperactivity which combined had 18 items. As the scale was developed in the US, the evaluation team in the previous trial discarded three items that were judged to be irrelevant for English classrooms. A copy of the scale is included in Appendix 1.

Baseline data from the SNAP-IV Teacher Attention Rating Scale will be used as a baseline for the same measure at endline. Given the ease of administration and the importance of controlling for prior attention levels at baseline, it was decided to include this measure at baseline. In the previous efficacy trial the pre- post-test correlation was 0.683.

## Primary outcome

The primary outcome for the effectiveness trial will be pupils' number skills as assessed by the number skills test that forms part of GL Assessment's British Ability Scales 3rd Edition (BAS3). The BAS3 is formed of 20 individual scales, of which number skills is one.

The BAS3 is a standardised battery of cognitive tasks that has long been established as a leading measure for assessing a child's cognitive ability and educational achievement (EEF, n.d.). It is widely administered by educational and clinical psychologists to assess children for learning and behavioural difficulties, and to identify both gifted and talented pupils and underachievers (GL Assessment, n.d.; EEF, n.d.). There is one version that is suitable for early years and another for school-aged children (Swinson 2013). For this assessment the school-aged version will be used. The battery involves a number of diagnostic and achievement scales covering Word Reading, Spelling and Number Skills (Swinson, 2013).

The BAS3 number skills achievement subscale will be administered on a one-to-one basis at post-test at the end of Year 3 (May 2022). The tests will be administered by independent test administrators from Alpha Plus who will be blinded to allocation status. While the test is designed to be used by psychologists, GL has agreed that it can be administered by appropriately trained individuals. Alpha Plus administrators will be trained in the use of BAS3 by a psychologist familiar with the test and its administration.

The number skills achievement subscale is designed for adaptive testing: children start the test at a level considered appropriate for their age and are presented with easier items if they do not meet a criterion for moving to more difficult items or are presented with progressively more difficult items until they can no longer meet the criterion for moving forward. The starting point for children aged seven to eight includes items that require the child to read or write 3-digit numbers and solve simple addition and subtraction word-problems presented orally (e.g. “Reuben had 18 crayons in his pencil case. He gave 5 to his friend. How many did he have left?”). It is expected that typically achieving children will meet the criterion to be presented with more difficult items, which are the starting point for the next age group. The more difficult items include arithmetic calculations with two-digit numbers without regrouping (e.g. 15+23) and with (e.g. 38+57) as well as word problems involving multiplication (e.g. “A set of batteries costs 2 pounds. How much do 4 sets cost?”). It must be noted that the IWM+A intervention does not include practice in arithmetic calculation or word problem solving; children are exposed to two- and three-digit numbers when they work on additive composition, but their task is neither to read nor to write the numbers. Thus, the BAS3 number skills subtest is not a measure of what was explicitly taught during the intervention.

Test administrators will be responsible for marking each question as the test is administered (in line with the overall BAS3 testing protocol), ensuring that marking will also be blind to group allocation. Total scores will be calculated by summing the number of questions answered correctly by pupils. Correct answers receive one point, while incorrect answers or no answers are scored as zero (Elaissen et al., 2017).

As the BAS3 uses Rasch Scaling, there is no requirement for all items to be administered in order for the results to be meaningful (Swinson, 2013). This means that it is appropriate to administer just one area of the test (e.g. number skills achievement subscale) according to the concerns being investigated (Swinson, 2013), as is the case here for number skills. The Rasch Scaling also means each scale can be accurately completed in five minutes (Swinson, 2013). The scales have been standardised on a sample of nearly 1500 British children from diverse geographical areas and ethnic backgrounds (Swinson, 2013).

The BAS3 number skills subscale was selected as it was used as the primary outcome in the efficacy trial, allowing for comparison across trials. We note that there were issues with the BAS3 number skills subscale in the efficacy trial: scores for each group were skewed, with a significant proportion of pupils scoring at or near the limit for the test (Wright et al., 2019). The movement of scores around a density of 0.3 across trial arms still indicated, however, that there was sensitivity in the measure for the top end of the distribution (Wright et al., 2019). To better understand the psychometrics of BAS3 for this particular population, TLs will administer BAS3 tests in Phase 1, with training in BAS3 number skills subscale delivered by Oxford University. Researchers from RAND Europe will then look at the results from the test to understand if concerns that arose in the efficacy trial are present. The BAS3 provides a variety of scores, including raw, age-standardised, and ability-scores. During Phase 1 piloting RAND will clarify which is the most appropriate score to use for the final evaluation, looking at distribution and range of scores. Further details will be provided in the Statistical Analysis Plan.

To minimise measurement attrition due to pupil absence, Alpha Plus will revisit schools where pupil absence due to illness is 20% or higher (i.e. two or more pupils missing on the day).

### Secondary outcomes

There are three secondary outcomes for this trial, including WM, wider maths attainment, and attention. These are outlined further below.

*Working memory*
Pupil's WM will be measured using the WM Test Battery for Children (WMTBC) (Pickering & Gathercole, 2001), using three sub-scales as per the efficacy trial: listening recall, counting recall and backwards digit recall. This measure was used in the previous efficacy trial. As outlined in the theory of change (see Figure 1), WM is a key proximal outcome that should be directly affected by the programme. This is supported by some of the findings from the efficacy trial that suggest some measurable impact can be detected.

In contrast to short-term memory tasks, which require simply repeating information after it was presented, working memory central executive tasks require processing information before an answer is given. The three types of information manipulation in the central executive tasks used in the WMTBC, namely comprehension, counting, and inverting the order of a list of items, are typical of all working memory tasks, and are also used in the intervention; the intervention also includes inhibition of irrelevant information. However, the way in which the information is processed in the intervention and in the WMTBC is not the same. For example, both include listening to a sentence, judging whether it is true or false, and then recalling the last word in the sentence. However, in the WMTBC the judgement is based on information that the child is expected to know (e.g. "All sisters are girls") whereas in the intervention the children listen to a sentence and judge whether it is true or false of a picture (e.g. The sentence "They went for a walk by the sea" is presented while the child looks at people walking on hills).

The WMB was designed by Gathercole & Pickering (2000a) in line with the Baddeley (2000) WM model. It is composed of 13 subtests in total, each of which reflects one of the three components of this model: phonological loop, visuo-spatial memory and central executive. The 10 subtests relating to phonological loop and visuo-spatial memory, with activities around forward digit recall, nonword repetition, serial recognition and recall of words and nonwords, as well as static and dynamic matrices and mazes, are not being used as measures for this trial (Gathercole & Pickering 2001). The three subtests used for this trial all relate to the central executive element of the Baddeley & Hitch model, which 'is believed to support a variety of activities including controlling the flow of information through WM, the retrieval of knowledge from long-term memory, the control of action, and the scheduling of multiple concurrent cognitive activities' (Gathercole & Pickering, 2000b).

In the **listening recall** subtest, the child listens to a series of sentences, judges the veracity of each sentence in turn, then recalls the final word of each of the sentences in the sequence (Gathercole & Pickerin,. 2001). The number of sentences in each trial then increases by one every four trials until the child incorrectly recalls two or more trials at the same sentence length (Gathercole & Pickering, 2001). The child's final score is the total number of trials correctly recalled (Gathercole & Pickering, 2001). Gathercole & Pickering (2000b) reported the test-retest reliability coefficient for this task to be 0.62.

In the **counting recall** subtest, the child counts aloud the number of dots in a series of arrays and is then asked to recall the dot totals in the order the arrays were presented (Gathercole & Pickering, 2001). The number of arrays is increased by one every four trials. The child's score is the total number of trials in which the number of dots were recalled in the correct order (Gathercole & Pickering, 2001).
Gathercole & Pickering (2000b) reported low test-retest reliability for this measure, at 0.15.

In the **backwards digit recall**, the child is presented with spoken sequences of digits that they are then asked to recall in reverse order. The test begins with four trials containing two digits, with longer sequences following on if three or more lists are correctly recalled. The child's score is the total number of lists correctly recalled (Gathercole & Pickering, 2001). This is a version of the Wechsler Intelligence Scale for Children (WISC-II), which includes both forwards and backwards recall of digits, and has a split-half of 0.85 (Golombok & Rust, 1992).

Gathercole & Pickering's assessment (2000b) of the battery's validity indicate high internal validity for the central executive measures, as well as good external validity, with the central executive tasks showing pervasive associations with achievements in the areas of language, literacy and mathematics.

The test will be administered at endline on a one-to-one basis by independent test administrators from Alpha Plus who will be blinded to allocation status. Given the high burden of test administration it was decided not to conduct the test at baseline. In the previous trial it was also administered at baseline, however, to reduce burden on schools there will be no administration of the WMTBC at baseline. SNAP-IV baseline scores will be used in the analysis as a baseline measure for working memory. Although the measure is based on teachers' reports and the correlation is relatively low, there is a relation between the constructs of attention and WM.

The subtests are combined by means of factor analysis to avoid content bias or process bias in the measure. The same approach will be used in this trial.

*Wider maths attainment*

Progress in maths is ultimately the main distal outcome for the IWM+A intervention and is of particular interest to the EEF (see Figure 1). To better understand the impact of IWM+A on maths attainment in general, a wider measure of maths attainment will be used to understand the extent to which IWM+A supports maths learning. Wider maths attainment will be measured using GL Assessment's Progress Test in Mathematics (PTM). PTM was selected as the wider maths attainment measure as it is well-known to schools and used in a number of historic and ongoing EEF trials. The raw scores will be used in the analysis. An additional analysis will be conducted using factor analysis to identify impact in the areas of the test that are most relevant to the intervention (see Additional analyses).

In consultation with GL it was decided that Version 8, designed to be used by Year 3 children in the spring/summer term, would be used instead of Version 7, despite the lower ability level of the pupils. It was felt that Version 8 was the most developmentally appropriate and that the underlying constructs and the range of items available would mitigate against potential floor effects.

The group-delivered test will be administered to all ten selected pupils in each school at the same time under exam conditions by independent test administrators from Alpha Plus who will be blinded to allocation status. Tests will be marked by GL markers who will not know school allocation, thus adding an extra layer of blinding. EYFSP will be used as a baseline.

*SNAP-IV Teacher Attention Rating Scale*

Attention and behaviour outcomes will be measured using the SNAP-IV Teacher Attention Rating Scale as per the efficacy trial. As outlined in the theory of change (see Figure 1), attention in class is considered to be an important proximal outcome leading to improved impact in maths. Further details on this measure can be found under 'Baseline measures'.

The scale needs to be completed by an individual (i.e. a teacher) who is familiar with the pupil and their behaviour, and as such will be completed by pupils' teachers, who will consequently not be blind to allocation. Teachers will be asked to complete these measures for all nominated pupils in May 2022.

## Compliance

For the purpose of this evaluation, we are employing the EEF's definition of compliance as 'the extent to which the critical ingredients of the intervention are delivered to and/or received by the target participants' (EEF, 2019b). In collaboration with the EEF and the delivery team, we have defined 'compliance' as pupil attendance at sessions. This will be a binary measure, defined at the pupil level, based on attendance as recorded by the TAs using a template developed by Oxford. Children who have been marked as having attended all five WM sessions and at least four arithmetic sessions will be marked as compliant; all those that do not obtain this will be marked as non-compliant. TAs will be asked to share the logs with RAND after all sessions have been completed.

## *Analysis*

### Primary outcome analysis

The primary analysis will be on an intention-to-treat (ITT) basis, under an analysed-as-randomised approach, whereby the analysis will include all randomised schools/pupils in the groups to which they were randomly assigned, regardless of the treatment actually received, withdrawal from the intervention post-randomisation, or deviations in programme implementation. The ITT approach is inherently conservative as it captures the averaged effect of offering the intervention, regardless of whether the participants complied with assignment. This principle is key in ensuring an unbiased analysis of intervention effects and is in line with the EEF's guidance (see EEF 2018).

The primary outcome will use standardised scores on BAS3, with prior attainment being accounted for by EYFSP scores (see the 'Outcome measures' section for more detail). To estimate the impact on the primary outcome we will use a three-level multilevel model (pupils in schools in regions) to account for clustering of data.

The main analysis consists of the model for outcomes of pupils nested in schools, which is:

(1) $Y_{ij} = \beta_0 + \mathrm{IWMA}_j\tau + Z_j\beta_1 + X_{ij}\beta_2 + u_j + e_{ij}$

where $Y_{ij}$ is the BAS3 score for child $_i$ in school $_j$; $\beta_0$ is the cluster-level coefficient for the slope of a predictor on number skills; $\mathrm{IWMA}_j$ is a binary indicator of the school assignment to intervention [1] or control [0]; $Z_j$ are school-level characteristics, here the stratifying variable of geographical location (as used for randomisation); $X_{ij}$ represents characteristics at pupil level (pupil $_i$ in school $_j$), specifically the pre-intervention EYFSP score; $u_j$ are school-level residuals and $e_{ij}$ are individual-level residuals.

Equation (1) is known as a 'random intercepts' model because $\beta_{0j}=\beta_0+u_j$ is interpreted as the school-specific intercept for school $j$ and $\beta_{0j} \sim_{i.i.d} N(\beta_0, \sigma u2)$ is random (it is a number that can take any value). Our target parameter (i.e. the focal result of the trial) is τ, a binary treatment/control indicator variable. The effect size (Hedge's *g*) will be standardised using unconditional variance in the denominator and confidence intervals will be reported to communicate statistical uncertainty in line with EEF guidance (see EEF 2018). This will tell us the average effect of the intervention on pupil outcomes in treatment schools compared to those in control schools.

### Secondary outcome analyses

The following secondary outcome analyses are planned: (1) WM, (2) wider maths attainment, (3) attention. This analysis will use the measures described in the 'Outcome measures' section.

Secondary outcomes will be assessed following the same specification to Equation (1) listed under 'Primary outcome analysis' above, but we will substitute either (1) the score on WMTBC or (2) scores on PTM8 or (3) scores on SNAP-IV for the secondary outcomes (i.e. *Y* in Equation 1).The vector of pupil-level characteristics in Equation (1), $Xij$, will include the relevant baseline scores from EYFSP (2) or SNAP-IV (3). It is not possible to include a direct baseline of WM as the administrative burden would be too high. Instead $Xij$ will include prior attainment using the EYFSP. Further details will be discussed in the Statistical Analysis Plan (SAP).

### Subgroup analysis

Although pupils eligible for free school meals (FSM) do not represent the focus group for this trial, this is an important subgroup, particularly given the findings from the previous efficacy trial (see 'Study rationale and background'). Analysis will be undertaken with the binary FSM variable (FSM-eligible=1; non-FSM-eligible=0) entered into the analysis (see Equation 1) as a moderator, therefore using the whole trial sample. Effect size calculations and statistical uncertainty will be calculated and communicated as per the primary outcome. More details will be provided in the SAP.

**Compliance**

As the ITT approach is inherently conservative, capturing the averaged effect of *offering* the intervention, we also propose to look at treatment effects in the presence of compliance at the school level as an additional analysis to capture the averaged effect of *participation* in the intervention.

In a situation of imperfect compliance, whereby not all participating schools are deemed compliant using the criteria outlined above, we will undertake a complier average causal effect (CACE) analysis, by drawing on an instrumental variable (IV) approach, and using a two-stage least squares (2SLS) estimation approach to recover the treatment effect for those who complied with assignment. Further details of how compliance will be analysed will be presented in the statistical analysis plan (SAP).

**Additional analysis**

We are also interested in understanding the role that 'dosage' has on outcomes, given the important role programme participation can have on variance in outcomes, particularly for younger children (Zhi et al., 2010). We therefore propose to undertake an additional analysis, using data from pupil attendance records as a continuous variable. This will be an additional analysis that will sit alongside the compliance measure to better understand how session attendance mediates outcomes. Further details will be provided in the SAP.

We also will use factor analysis to understand the extent to which specific factors in PtM linked to IWM+A are impacted by the intervention. Further details will be provided in the SAP.

**Imbalance at baseline**

A well-conducted randomisation should yield groups that are equivalent at baseline, with any imbalance at baseline occurring by chance (Glennerster & Takavarasha 2013). However, to check for, and monitor, imbalance at baseline in the realised randomisation, baseline equivalence testing will be conducted at the school and pupil level.

At the school level, we will check the balance in the following variables by means of cross-tabulations and histograms that assess the distribution of each characteristic within the control and intervention groups:

- Ofsted ratings
- Proportion of children eligible for FSM

At the pupil level, balance will be assessed as above but for the following characteristics:

- Gender
- Prior attainment using the EYFSP
- FSM status for pupils

**Missing data**

Missing data can arise from item non-response or attrition of participants at school and pupil levels. Even though it is important to include all data, it can be problematic to apply the intention to treat principle if we are not able to complete follow up testing for all randomised schools. To better understand the pattern of missing data and its impact on the analysis, we will explore the extent of missingness, and whether there is a pattern in missingness.

We will analyse and report missingness for the primary outcome measure (and associated primary analysis); for the secondary outcome measures and FSM subgroup we will report the extent of missing data but not undertaken any additional analyses. The procedure will be outlined in detail in the SAP.

### *Longitudinal follow-ups*

Longitudinal analysis will be commissioned separately to this trial.

## Implementation and process evaluation

Implementation is the process by which an intervention is put into practice (Lendrum & Humphrey, 2012) and can be considered as a multi-dimensional construct, consisting of compliance, fidelity, participant responsiveness, programme differentiation, monitoring of control/comparison conditions, and adaptation (Humphrey et al., 2016).

### *Research questions*

The process evaluation will address the following questions:

RQ1. To what extent was IWM+A delivered as intended?

- a) What components of the different aspects of the intervention were delivered with the highest fidelity (and why)? Which were implemented with less fidelity (and why)?
- b) What adaptations were made (and why) and what (if any) impact did they have on pupil responsiveness?
- c) To what extent are there differences in adaptations and pupil responsiveness between FSM and non-FSM pupils?

RQ2. What appear to be the necessary conditions for success? What appear to be the barriers to delivery?

RQ3. To what extent is the 'train the trainer' model a viable approach for delivering the intervention at a larger scale? What changes could be made (if any) to improve delivery at scale?

RQ4. What was 'business as usual' in the control schools?

RQ5. To what extent is the theory of change a valid representation of IWM+A as delivered over the course of the trial?

- a) In what way does the broader context influence delivery?
- b) To what extent were the predicted moderators (e.g. training quality, preparation time, space for delivery, attendance at training, senior leadership 'buy-in') a factor during implementation? Were other potential moderators identified that were not predicted? [3]

### *Fidelity (i.e. core components)* [4]

As part of an IDEA workshop the delivery team outlined the following elements that were considered to be core components (i.e. definition of fidelity):

- TA attendance at all training sessions
- Child and TA attendance at five WM sessions and at least four arithmetic sessions
- Appropriate learning space (i.e. quiet, free from interruptions)
- Each child has computer
- TAs scaffolding the use of strategies when required, giving positive feedback on the use of strategies rather than on the correct responses, attributing success to the use of strategies
- Variation in games (i.e. children play a number of different games per session)
- Games increase in difficulty
- Participation both in TA-led and independent sessions

[3] For the purpose of this evaluation, we are employing the EEF's definition of moderators as 'variables that modify the form or strength of the relation between intervention and outcome' (EEF 2019a). Moderators can be individual characteristics or contextual factors and address the question: 'For whom does the intervention work and under what circumstances?' (EEF 2019a).

[4] For the purpose of this evaluation, we are employing the EEF's definition of fidelity as 'the degree to which the intervention is delivered as intended or prescribed' (EEF 2019a).

These will be collected via TA attendance logs (to address the first two bullet points) while the rest of the fidelity aspects will be collected via observations and survey (see below).

### *Research methods*

We have developed a mixed-methods Implementation and Process Evaluation (IPE) data collection plan which includes document reviews, semi-structured focus group discussions, observations, surveys and semi-structured interviews, as outlined in Table 3. The tools and approaches are based on the theory of change which was co-constructed with the delivery team and the EEF. This ensures that the approach is theory-based and intervention-led. Finally, we have balanced the need to triangulate data (i.e. to increase reliability of findings by asking a number of sources) with the need to reduce burden on school staff.

**Table 3: IPE methods overview**

| Research methods | Data collection methods | Participants/ data sources (type, number) | Data analysis methods | Research questions addressed | Implementation/ theory of change relevance |
|---|---|---|---|---|---|
| Document Analysis | Training logs | TA training attendance logs, developed by Oxford and collected by TLs (12) | descriptive statistics | RQ1 | Fidelity |
| | Session logs | Session attendance logs (i.e. pupil attendance), developed by Oxford and collected by TAs (120) | descriptive statistics, IV approach as part of compliance | RQ1 | Compliance |
| Observations | Training Observations (RAND) | WM session delivered by Oxford to TLs in March 2021 (1); WM and Arithmetic training-to-train sessions delivered by Oxford to TLs in 2021/2022 (2). | Thematic analysis | RQ1, RQ2, RQ3 | Fidelity, Moderator |
| | Training Observations (RAND ) | Sample of training sessions delivered by TLs to TAs (8 total: 4 WM and 4 Arithmetic) | Thematic analysis | RQ1, RQ2, RQ3 | Fidelity, Moderator, Adaptations |
| | Structured Session Observations by | All TLs (12); Random sampling of | descriptive statistics, | RQ1, RQ2, RQ3, RQ5 | Fidelity, Adaptation, Pupil Responsiveness[5] |

[5] Participant responsiveness refers to the extent to which the intended recipients of an intervention are engaged by the activities and content that is delivered during implementation (Dusenbury et al., 2003).

| | | | | | |
|---|---|---|---|---|---|
| | RAND (using observation tool developed by RAND) | sessions from TAs in 20 case study schools; FSM eligibility will be indicated by the school allow deeper understanding of any differential effects | Thematic analysis | | |
| | Structured Session Observations by TLs and Oxford (using observation tool developed by Oxford) | All TLs (12); random sampling of TL and Oxford observations from TAs in case study schools (20) | descriptive statistics, thematic analysis | RQ1, RQ2, RQ3 | Fidelity |
| Focus Groups | Semi-structured interviews (RAND) | TLs (12) | Thematic analysis | RQ3 | Fidelity, Context |
| Interviews | Semi-structured interviews (RAND) | TAs in case study schools (20) | Thematic analysis; IV approach as part of compliance (for some items) | RQ1, RQ2, RQ3 | Fidelity, Adaptation, Context, Moderators |
| Survey | Online questionnaires (RAND) | Headteachers and TAs in intervention schools (120) | Thematic analysis, descriptive statistics, | RQ1, RQ2, RQ3 | Fidelity, Adaptation, Moderators, Costs |
| Survey | Online questionnaires (RAND) | Headteachers, Teachers and TAs in control schools (120) | Thematic analysis, descriptive statistics, | RQ4 | Business as usual |

A key underlying principle of the IPE measures is to understand how TL training and preparation influences the way in which IWM+A training is delivered to TAs and how this in turn influences programme delivery at the TA-pupil level. To do this, we will explore how successfully (or not) the core components of IWM+A are transferred from Oxford University trainers to IWM+A TLs in Phase 1 and from IWM+A TLs to TAs in Phase 2.

Wherever possible, measures developed and used in Phase 1 will also be used during the trial in order to map across the various phases of implementation. Further details on each method are provided below.

1. *Document Review*

Several documents are being collected by the delivery team (e.g. Oxford University and TLs) that will be used by RAND Europe to understand implementation. These documents will be used as they reduce data collection burden on TAs.

- **TA training attendance logs**: these will be collected by TLs at the training and will be a record of TAs that attend training. These will be collected and used by RAND Europe to understand fidelity (i.e. 'the degree to which the intervention is delivered as intended or prescribed).
- **Session attendance logs**: a record of pupil attendance will be collected by TAs following a template developed by Oxford. All TAs will be asked to complete these and share these with

the RAND team once all sessions have been completed. Data from these logs will be used to understand compliance (see ‘Compliance’ section).

2. *Observations*

*Observations of training*

Observations of the training will be carried out by RAND Europe to understand the way in which training imparts the core components of the intervention and how these are cascaded from Oxford to TLs to TAs (and eventually to programme delivery). Observers from RAND will attend the WM training session delivered by Oxford to TLs in March 2021, as well as two training-for-training days delivered by OU to TLs in 2021/2022. Researchers will also attend eight training sessions delivered by TLs to TAs in a selected sample of schools (see ’deep dive’ schools), with the aim to balance observations across WM and arithmetic training (i.e. four observations for each training).

RAND researchers will observe the training with the aim of understanding how training is cascaded from Oxford to TLs and from TLs to TAs. This will shed light on the extent to which training imparts core elements of programme to those responsible for delivering the programme. This will help to understand the extent to which the train the trainer model is working (or not working) in practice. RAND researchers will make notes on key features of the training, including content, structure, and timing. Notes will be made during training sessions. These notes will be analysed using a core components framework that will allow researchers to map key details of training against the core components. Findings will be compared across TL training and TA training to understand the similarities and difference between the different trainings. These will be summarised and compared to session observations (see below).

*Structured Session Observations by RAND (using observation tool developed by RAND)*

The aim of RAND’s observations is twofold. The first is to record fidelity of intervention delivery, pupil responsiveness, and adaptations. This will help us understand how the intervention is delivered and the extent to which moderators – in this case pupil responsiveness and adaptations – are present and how they interact with programme delivery. The second is to observe the extent to which core elements of training are observed and any patterns or differences that emerge during delivery. This will help us understand and describe the process of transmission within the train-the-trainer model.

During Phase 1, a remote observation tool was developed and piloted by RAND. The tool captures a number of elements considered to be central to delivery (i.e. fidelity), including the explicit teaching and use of strategies in sessions as well as pupil responsiveness, and notes any adaptations in delivery. The observation tool was designed to produce a mix of quantitative and qualitative items. One element of the tool – the matrix of teaching and scaffolding strategies – was developed by Oxford for their observation tool (see below). The other items were developed by RAND. The tool has been developed so that it can be used in-person or with remote observation sessions. Remote observations were conducted in Phase 1 and a mix of in-person and remote observations will be used in Phase 2. This flexibility of remote and in-person usability is required to mitigate risks of societal restrictions due to the COVID-19 pandemic and ensures that a number of schools can be observed within the available resources.

In Phase 1, RAND observed TLs delivering one WM sessions and one arithmetic session to pupils. These were conducted at the same time as Oxford observations. During Phase 2, RAND will conduct independent observations using RAND’s observation tool. RAND Europe will observe at two time points in each of the 20 deep dive schools (see ‘Deep dives’’ for further details) – once during the WM delivery, and once during the arithmetic delivery. RAND researchers will request to be present at the school on one day, either remotely or in person, and will observe all sessions taking place during that day. RAND is keen to observe more than one session per visit as Phase 1 focus groups suggested that delivery, engagement, and session content varied depending on the pupils involved. Observing a number of sessions over the course of one day is the best way to ensure variety in implementation can be observed. This also reduces the amount of visits that need to be arranged – ensuring that burden on schools is kept to a minimum. Visits (remote or in-person) will be scheduled so they do not take place at the same time as TL observations.

The observation tool will be analysed using the same core components framework developed for the analysis of training, allowing researchers to map key details of training against the core components. This will allow researchers to understand the extent to which core elements of training were observed in delivery and any patterns or differences that emerge during delivery. In addition, adaptations and other qualitative elements will be reviewed using a thematic analysis framework to identify and describe emerging themes. Quantitative items in the observation tool will be analysed to understand differences between different deep dive schools.

*Structured Session Observations by TLs and Oxford (using observation tool developed by Oxford)*

During Phase 1 Oxford developed an observation tool for the purposes of supporting quality assurance and fidelity assessment, and to guide professional learning conversations. In Phase 1, RAND used Oxford's observation tool with the aim of understanding how it might be used in the independent observation and to support moderation. At the same time, RAND was developing their own tool (see above).

During Phase 2 TLs and Oxford will also be observing sessions using an updated observation tool. The TLs will use this as a way of providing support and development to the TA; Oxford will use the tool to support fidelity and quality assurance in Phase 2, as well as support TL delivery. TLs will be trained on use of the tool in the September training day.

RAND Europe will collect TL and Oxford's observations in the deep dive schools with the aim of understanding fidelity. TL observations will be reviewed against the core components framework allowing researchers to map key details of TL observations against the core components and understand the extent to which they were observed by TLs. These will be combined with the key details from the TA training to see if patterns emerge when training core components and implementation core components are compared. These will be triangulated with interviews (see below) to understand if any patterns emerge that help us understand variation in fidelity and their causes.

Where possible, we would also like to review Oxford's observations of sessions. These will follow the same analysis approach as TL observations, with key details mapped against the core components framework. The analysis will be used to understand the extent to which Oxford's processes supported fidelity, and to understand better any barriers or facilitators to this aspect of scale-up (i.e. quality assurance, supporting TLs deliver at scale).

3. *Focus Groups*

RAND Europe will conduct four semi-structured focus groups with the eleven independent TLs[6]. These will happen: (1) soon after their WM and arithmetic training sessions in Phase 1; (2) near the end of their delivery of the arithmetic sessions in Phase 1; (3) soon after their WM training-for-training session at the start of Phase 2; and (4) near the end of the TA's delivery of the arithmetic sessions in Phase 2. The aim of the first two focus groups will be to gauge their understanding of IWM+A, of their role within the intervention, and their experiences of delivering IWM+A to children in Phase 1 schools. The final two focus groups will be used to understand the TLs' perspectives of acting as trainers and how they were supported (or not). A list of key questions will be drawn up and used following a similar approach to the semi-structured interviews.

[6] There are 12 TLs in the trial. One of the 12 TLs is employed by Oxford and will therefore not be part of the focus groups to avoid influencing comments from the other TLs.

4. *Interviews*

In Phase 2 RAND will conduct semi-structured interviews with TAs in 20 deep dive study schools at two time points to gain TAs' perspectives of delivering the programme as well as their thoughts on the training and support they received. These will be organised to coincide with the observations to minimise burden on TAs. This means that TAs will be interviewed once regarding the WM elements and once on the Arithmetic elements. If possible, RAND will also speak to Link Teachers in each case study school during the Arithmetic observations.

5. *Surveys*

A baseline survey exploring usual practice (i.e. use of targeted WM or math support) and perceptions of the IWM+A programme and WM in general will be sent after randomisation using an online survey tool. Surveys will be conducted after randomisation as the timeline prior to randomisation will not allow sufficient time for the survey to be conducted alongside all other pre-randomisation requirements.

An endline survey covering topics such as potential unintended consequences of implementing IWM+A, training, delivery, fidelity, and adaptations will be carried out with intervention school headteachers and TAs. Headteachers will also be asked about the cost implications of running the programme.

Control school staff (headteachers and TAs) will also be surveyed at endline to describe business as usual in the control groups. Survey questions will explore the extent to which control schools have implemented targeted maths and/or working memory interventions, including their cost.

*Deep dives*

To support a better understanding of how IWM+A is implemented in practice and to better understand scale-up a number of schools will be selected to act as sources of evidence in a series of 'deep dives'. One school will be selected from each region, selected to represent a balanced sample across key factors that might influence implementation. First, we will rank schools by average FSM-rating and aim to disproportionately sample from these schools. This will help us understand potential differential effects including how pupil response and/or delivery are moderated by pupil FSM eligibility. We will then aim to select 20 schools from a range of school sizes and Ofsted ratings.

Interviews and observations will be used within and across settings to provide vignettes that will describe how the intervention is working in practice, and to support conclusions that cut across the deep dive data collection activities (i.e. using interviews to make sense of observations).

## Analysis

Analysis will be carried out through a combination of simple quantitative analysis (i.e. descriptive statistics), an instrumental variable approach, and qualitative analysis using data from the research methods described above and in Table 3. For example:

- A simple quantitative analysis will provide **descriptive statistics** (e.g. total, average, range) to summarise responses from the attendance log, surveys, and structured observations. This will provide an insight into the extent to which different views are representative of the population involved in the trial. For example, results from the attendance logs will help describe the extent to which TAs attended training across the regions and whether there are any differences between regions. We will also look across phases on relevant items from measures, such as the core component framework and observation tools to understand the extent to which training in Phase 1 supports training and delivery in Phase 2.
- An **inductive approach** will be used to code qualitative data from interviews, focus groups and any open ended questions from surveys. To minimise bias the approach to interrogating data will be based on key elements of the theory of change (e.g., training, session core components, moderators, context), and will also factor and include key factors identified in the implementation/theory of change in Table 3. Credibility will be enhanced by triangulation across

data collection approaches (i.e. interviews, surveys, observations) and researcher triangulation (i.e. more than one researcher involved in data collection and analysis)
- **Thematic analysis** will be undertaken to analyse data from the deductive coding using themes that will support us to understand and answer the research questions. For example coding pertaining to fidelity will be analysed to understand the extent to which IWM+A is delivered with fidelity, as well as identify any implementation factors that might act as barriers or facilitators.

Quotes (if used) will be selected based on their ability to succinctly summarise prevalent key themes emerging from the analysis. To minimise bias, quotes will only be sought once key themes have been identified and will be agreed upon by all members of the report writing team before inclusion in the final report.

## Cost evaluation

Costs will be evaluated using data gathered through the interviews and surveys administered to headteachers from across all schools. We will establish the counterfactual by evaluating the cost of business as usual in control schools, which may include the direct and indirect costs of running programmes comparable to IWM+A. We will also seek to understand costs associated with scale up (i.e. Phase 1 costs) by asking Oxford University to provide information.

There are two levels of cost evaluations for this train-the-trainer trial: costs incurred in Phase 1 (training the TLs), costs incurred in Phase 2 (where TLs train TAs and TAs deliver IWM+A). Costs in Phase 2 are incurred by the participating schools, through direct, indirect, and marginal costs. Costs in Phase 1 are incurred by Oxford University as a result of scaling up the programme. The cost of IWM+A implementation will be calculated per pupil-school-years (i.e. costs per pupil over the course of three years) to facilitate comparison with other programmes, as per EEF guidelines (EEF 2019a). We will also provide costs for Phase 1 as a means of understanding the costs associated with this particular scale-up approach.

The aspects of costs incurred by schools that will be gathered through our data collection tools include: direct costs of running the programme, teacher time used for training, preparation, and delivery of the programme, supplemental material cost incurred to deliver the programme, additional staff time used to support TAs delivering the programme (e.g. the need to employ other TAs to do tasks that the IWM+A TA would have done had they not been delivering the intervention). Costs such as time spent, stationery and other supplies will be monetised using market estimates and we will use sensitivity analysis to account for heterogeneity of costs between schools. The evaluation will measure pre-requisite costs (e.g. presence of computers), start-up costs (e.g. time spent on training), and recurring costs (e.g. costs of materials, staff time required for support). Since delivery of IWM+A relies heavily on the use of computers, and since schools may be using computers for other purposes and for time periods beyond the delivery of IWM+A, our cost calculations will take into account these overlapping uses and life use of computers so as not to inflate incorrectly the costs of running the programme. Interviews with the TAs will also provide insight on hidden costs or savings from the programme that cannot be identified through surveys with headteachers.

## Ethics and registration

The trial will be registered on the International Standard Randomised Controlled Trial Number (ISRCTN) registry, which is used to describe randomised controlled trials (RCTs) and efficacy trials at inception. Once registered, this protocol will be updated with the assigned registration number.

The ethics and registration processes are in accordance with the ethics policies adopted by RAND Europe and Oxford University. The evaluation is approved by both the RAND U.S. Human Subjects Protection Committee (HSPC) and the University of Oxford Central University Research Ethics Committee (CUREC).

Parents or legal guardians act as decision-makers for individual pupils. This is because the intervention will be delivered during the school day, where schools act *in loco parentis*. Prior to pupil data being sent to the delivery team, parents will be sent information sheets and withdrawal forms by the school and parents will have the opportunity to return these. Parents can withdraw their children at any time from the data collection activities.

If participants choose to withdraw their children from the study later on, their data will not be collected or will be deleted, as appropriate (see the privacy notice in Appendix 2).

RAND Europe will collect consent forms for all TAs and TLs that participate in an interview or focus group. The front page for each online survey will contain a privacy notice informing respondents that participation in the survey is entirely voluntary. The consent form in the survey will be built into the data collection tool so that those moving past a certain page (following the privacy notice and information on the research) will have given consent for the data to be used in the research.

None of the evaluation team has any conflicts of interest and all members of the study team have approved this protocol prior to publication.

# Data protection

### *Legal basis for processing data*

Our team has extensive experience handling personal data, and our researchers are accredited by the Office for National Statistics to use, for instance, data from the National Pupil Database. RAND Europe would obtain personal data from schools and pupils as a data controller. The data collected for the IWM+A project is used for the purposes of research. The lawful basis for RAND Europe's use of that data under the General Data Protection Regulation (GDPR) is 'legitimate interest'.[7] Legitimate interest is an appropriate basis because the data collected as part of this evaluation will be used in ways that people would reasonably expect (i.e. for the benefit of improving provision for children in need of math support) and that have minimal privacy impact. Legitimate interests apply where processing is necessary for the purpose of the legitimate interest pursued by the controller (see GDPR Article 6 (1) (f)) and for statistical and research purposes (See GDPR Article 89). The University of Oxford (the delivery team) are also relying on legitimate interest as a legal basis.

### *Data protection*

The evaluation team will use basic identifiers (name and date of birth) to associate information with individuals to create datasets for research purposes. The rights and freedoms of the subjects will not be affected as information will only be identifiable during processing to the evaluation and delivery teams and not otherwise. If parents choose to withdraw their children from this study, their children's data will not be collected, or will be deleted if already collected. Research data (not sensitive/personal data) will be kept securely by the evaluation and delivery teams for the duration of the study and deleted one year thereafter.

The evaluation team have put appropriate security measures in place to keep personal data secure and to prevent any unauthorised access to or use of it. The evaluation team will collect and store all evaluation data in accordance with the Data Protection Act (2018) and GDPR requirements. Evaluation data will be stored on secure servers. Data transferred between the delivery and evaluation teams will be encrypted or use secure file transfer protocols. Data will be shared securely using specialised software (Syncplicity). No data will be saved on servers or shared with processors outside the United Kingdom or the European Economic Area (EEA) or the United Kingdom (if outside the EEA) pursuant to EEA approved terms.

[7] For more information about legitimate interest, please see: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/legitimate-interests/what-is-the-legitimate-interests-basis/.

## *How data will be used*

The table below describes all the ways that children's personal data will be used by the delivery team and the evaluation team.

| **Personal Information Data** | | | | |
|---|---|---|---|---|
| **Data**<br><br>**(what will be collected?)** | **Source**<br><br>**(who will collect it?)** | **Receiver or permitted recipient** | **Purpose**<br><br>**(what will be used for?)** | **How do we collect the data?** |
| **Pupil data provided by the schools (data already held by schools):**<br><br>Administrative pupil data for Year 3 pupils participating in the study to be requested from schools:<br>- First name, surname, Unique Pupil Number (UPN), Free School Meal status (FSM), Date of birth, Gender | Oxford University (OU) (data controller) will collect this from schools to allow for matching pupil data from the National Pupil database<br><br>(Data controller) | OU and RAND Europe<br><br>(Data controllers) | Matching pupil records; independent assessment of impact of intervention on outcomes; withdrawal forms; academic publications | Electronic data transfer |
| | | EEF<br><br>(Data controller)<br><br>and EEF archive manager (data processor) | Long-term follow-up analyses and archiving of the data. EEF becomes data controller once the data has been transferred to the EEF Data Archive, but is not data controller until this has happened. | |
| **Pupil data from the National Pupil Database:**<br><br>- Free School Meal status (FSM), Attainment data from the Early Years Foundation Stage Profile | National Pupil Database<br><br>(Data controller) | OU and RAND Europe (data Controllers) | Independent assessment of impact of intervention on outcomes; longitudinal analysis of children's development; academic publications | Electronic data transfer |
| | | EEF<br><br>(Data controller) and EEF archive (data processor) | Long-term follow-up analyses and archiving of the data. EEF becomes data controller once the data has been transferred to the EEF Data Archive, but is not data controller until this has happened. | |
| **Pupil outcome data collected for the evaluation (collected by test administrators):**<br><br>- BAS3 test –subtest Number Skills<br>- GL PTM8<br>- Working Memory Battery for Children – subtests Central Executive | RAND Europe<br><br>(Data controller)<br><br>(Data collection sub-contracted to Alpha Plus to test provider as Data Processor) | OU and RAND Europe<br><br>(Data controllers)<br><br>Alpha Plus<br><br>(Data processor) | Independent assessment of intervention impact on number skills and pupil's WM, academic publications | Paper forms sent by secure mail; Electronic transfer via secure server |
| | | EEF | Long-term follow-up analyses. EEF becomes data controller once the data has been transferred to | |

| | | | | |
|---|---|---|---|---|
| | | (Data controller)<br><br>and EEF archive (Data processor) | the EEF Data Archive, but is not data controller until this has happened. | |
| **Pre-test and outcome data collected for the evaluation (collected by schools):**<br><br>- Abbreviated version of SNAP-IV (Teacher rating scale of attention) | OU and Schools (Data processor)<br><br>OU and RAND Europe<br><br>(Data controller) | OU and RAND Europe<br><br>(Data controllers) | Independent assessment of impact of intervention on outcomes; longitudinal analysis of children's development; academic publications | Electronic transfer via secure server |
| | | EEF<br><br>(Data controller)<br><br>and EEF archive (Data processor) | Long-term follow-up analyses. EEF becomes data controller once the data has been transferred to the EEF Data Archive, but is not data controller until this has happened. | Electronic data transfer via secure server |
| **Data on delivery of IWM+A**<br><br>- IWM+A games usage data (i.e. number of games played; number of games with 100% correct answers - collected by the delivery team through the game software)<br>- Highest level of game achieved with TA and online; paper data collected from schools<br>- Delivery team quality assurance records of TAs delivery of IWM+A sessions (children are present but no child data will be collected) | Schools (Data processor)<br><br>University of Oxford (Data controller) | OU and RAND Europe<br><br>(Data controllers) | Analyses of quality assurance and dose of intervention; academic publications | Electronic data transfer via secure server |
| | | EEF<br><br>(Data controller)<br><br>and EEF archive (Data processor) | Long-term follow-up analyses. EEF becomes data controller once the data has been transferred to the EEF Data Archive, but is not data controller until this has happened. | Electronic data transfer via secure server |
| **Contact information**<br><br>- **Emails and names of TAs and Link Teachers** | Schools (Data processor)<br><br>University of Oxford (Data processor) | OU and RAND Europe (Data controller) | To maintain contact with schools and to send electronic questionnaires | Electronic data transfer via secure server |

## Personnel

***Delivery team: University of Oxford***

Project Leaders: Gabriel Stylianides & Terezinha Nunes

Research Project Manager: Louise Matthews

Research Officer: Rosanna Lea

***Evaluation team: RAND Europe***
Principal Investigator and overall project leader: Elena Rosa Brown
Project Manager: Lillian Flemons
Fieldwork and analysis team: Andreas Culora, Emma Leenders, Lydia Lymperis

## Risks

| Risk | Assessment (likelihood/impact) | Mitigation strategy |
|---|---|---|
| **Recruitment failure** | Likelihood: Moderate<br>Impact: High | This can be mitigated by regular dialogue over any recruitment issues. |
| **Attrition** | Likelihood: Moderate<br>Impact: High | We propose recruiting more schools to build in a 'buffer' for attrition at person and setting level. Schools and parents of pupils selected in the study are given clear information about participation before signing up. |
| **Low implementation fidelity** | Likelihood: Moderate<br>Impact: Moderate | Triangulated process evaluation across Phase 1 and 2 to monitor and document fidelity of implementation. |
| **Low participation rates for IPE surveys and interviews** | Likelihood: Moderate<br>Impact: Moderate | Sufficient data collection window given with real-time monitoring of response rates to allow for reminders to be targeted. |
| **Cross-contamination** | Likelihood: Low<br>Impact: High | Clear inclusion criteria to stress no use of targeted mathematics interventions with targeted pupils; information about other comparable programmes will be factored into the analysis. |
| **Quality of reporting** | Likelihood: Moderate<br>Impact: Moderate | Applying RAND QA processes including expert review. |
| **Lack of coordination between RAND Europe, the EEF and the delivery team** | Likelihood: Moderate<br>Impact: Moderate | Teams to attend initial meetings and agree on roles and responsibilities at the outset.<br>Regular contact between senior team from each organisation. |
| **Evaluation team members absence or turn-over** | Likelihood: Moderate<br>Impact: Low | All RAND Europe staff have a three-month notice period to allow sufficient time for handover.<br>The team can be supplemented by researchers with experience in evaluation from the larger RAND Europe pool. |
| **Data security breach** | Likelihood: Low<br>Impact: High | Data sharing agreement is in place between RAND Europe and Oxford University and a protocol to act on in the case of a breach. |

| Risk | Assessment (likelihood/impact) | Mitigation strategy |
|---|---|---|
| **Disruptions to data collection activities due to COVID-19** | Likelihood: Moderate<br>Impact: High | We have built in a range of remote data collection approaches to ensure disruptions to data collection are kept at a minimum; we will also be sure to revisit timelines if needed to ensure data collection does not coincide with particularly stressful time for schools. |

## Timeline

The timeline for the evaluation can be seen in Table 4. It should be noted that there was a significant delay between the second set-up meeting and TL recruitment and the TL training and practice. This is due to disruptions caused by COVID-19 which closed schools to all but essential workers from March 2020. As such, all evaluation and delivery activities were put on hold. These were reactivated in January 2021 with the IDEA workshop.

**Table 4: Timeline**

**PHASE 1**

| Dates | Activity | Staff responsible/ leading |
|---|---|---|
| **Jan 2020** | First set-up meeting | EEF |
| **Feb 2020** | Second set-up meeting | EEF |
| Oct – Dec 2020 | Recruiting TLs | Oxford University |
| **Jan 2021** | IDEA workshop | RAND Europe |
| Jan – June 2021 | TLs attend training and practice delivering to pupils | Oxford University |
| April – June 2021 | Oxford and RAND observe TL delivery to pupils | Oxford University, RAND Europe |
| **May 2021** | Second IDEA workshop | RAND Europe |

**PHASE 2**

| Dates | Activity | Staff responsible/ leading |
|---|---|---|
| Jan – May 2021 | TLs recruit schools | Teacher Leaders |
| Sept 2021 | Pupil nomination and baseline data collection | Schools, Oxford University, RAND Europe |
| Sept 2021 | Randomisation during the final week of September | RAND Europe |
| Dec 2021 | Application to access NPD data | RAND Europe |
| Sept 2021 | Training for TLs in how to manage delivery of WM to TAs | Oxford University |
| Oct 2021 | Training of TAs in WM | Teacher Leaders |

| | | |
|---|---|---|
| Oct – Dec 2021 | TAs deliver WM sessions | Schools |
| Jan 2022 | Training for TLs in how to manage delivery of arithmetic to TAs | Oxford University |
| Jan 2022 | Training of TAs in arithmetic | Teacher Leaders |
| Dec – Feb 2022 | TAs deliver arithmetic sessions | Schools |
| Nov - February | Observe TA delivery to pupils (as part of delivery) | Oxford University, Teacher Leaders |
| Nov - February | Observe TA delivery to pupils (as part of evaluation) | RAND Europe |
| April 2022 | Surveys | RAND Europe |
| May 2022 | Outcome testing | RAND Europe |
| Sep 2022 | First Draft EEF report | RAND Europe |
| Mar 2023 | Final EEF report, submission of data to the EEF data archive and updating the ISRCTN trial registry with results | RAND Europe |

## References


Baddeley, A. & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), *Recent Advances in Learning and Motivation,* vol. 8, 47-90, New York: Academic Press.

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Science, 4*(11), 417-423.

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63*, 1-29.

Barahmand, U. (2008). Arithmetic disabilities: training in attention and memory enhances arithmetic ability. *Research Journal of Biological Sciences*, *3*(11), 1305-1312.

Bussing, R., Fernandez, M., Harwood, M., Hou, W., Garvan, C. W., Eyberg, S. M., & Swanson, J. M. (2008). Parent and teacher SNAP-IV ratings of attention deficit hyperactivity disorder symptoms: psychometric properties and normative ratings from a school district sample. *Assessment, 15*(3), 317-328.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin, 104*(2), 163.

Cowan, N. (2014). Working memory underpins cognitive development, learning, and education. *Educational psychology review,* 26(2), 197-223.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior, 11*(6), 671-684.

De Smedt, B., Janssen, R., Bouwens, K., Verschaffel, L., Boets, B., & Ghesquière, P. (2009). Working memory and individual differences in mathematics achievement: A longitudinal study from first grade to second grade. *Journal of experimental child psychology*, *103*(2), 186-201.

Dong, N. & Maynard, R. (2013). PowerUp!: A tool for calculating minimum detectable effect sizes and minimum required sample sizes for experimental and quasi-experimental design studies. *Journal of Research on Educational Effectiveness,* 6(1), 24-67.

Duncan, J. (2006). Brain mechanisms of attention. *The Quarterly Journal of Experimental Psychology, 59*, 2-27.

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. *Health Education Research*, *18*, 237–56.

EEF. (2018). *Statistical analysis guidance for EEF evaluations*. Education Endowment Foundation. Retrieved 17 August 2021, from: https://educationendowmentfoundation.org.uk/public/files/Grantee_guide_and_EEF_policies/Evaluation/Writing_a_Protocol_or_SAP/EEF_statistical_analysis_guidance_2018.pdf

EEF. (2019a). *Cost Evaluation guidance for EEF evaluators*. Education Endowment Foundation. Retrieved 17 August 2021, from: https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting_up_an_Evaluation/Cost_Evaluation_Guidance_2019.12.11.pdf

EEF. (2019b). *Implementation and process evaluation guidance for EEF evaluations.* Education Endowment Foundation. Retrieved 17 August 2021, from: https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting_up_an_Evaluation/IPE_guidance.pdf

EEF. (n.d.). *British Ability Scales.* Education Endowment Foundation. Retrieved 17 August 2021, from: https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluating-projects/early-years-measure-database/early-years-measures-database/british-ability-scales/

Elaissen, E. et al. (2017). Is cognitive development at three years of age associated with ECEC quality in Norway? *European Early Childhood Education Research Journal,* 26(1), 97-110.

Friso-van den Bos, I., van der Ven, S. H. G., Kroesbergen, E.H. & van Luit, J. E. H. (2013). Working memory and mathematics in primary school children: A meta-analysis. *Educational Research review*, *10,* 29-44.

Gathercole, S. E. & Pickering, S. J. (2000a). Assessment of working memory in six and seven year old children. *Journal of Educational Psychology, 92*, 377-390.

Gathercole, S. E. & Pickering, S.J. (2001). Research Section: Working memory deficits in children with special educational needs. *British Journal of Special Education, 28*(2), 89-97.

Gathercole, S. E. & Pickering, S. J. (2000b). Working memory deficits in children with low achievements in the national curriculum at 7 years of age. *British Journal of Educational Psychology,* 70(177).

GL Assessment. (n.d.). *BAS3.* GL Assessment. As of 17 August 2021, from: https://www.gl-assessment.co.uk/assessments/products/bas3/

Glennerster, R. & Takavarasha, K. (2013) *Running Randomized Evaluations: A Practical Guide*. London: Princeton University Press.

Golombok, S. & Rust, J. (1992) *Manual of the Wechsler Intelligence Scale for Children, Third Edition UK.* Kent: The Psychological Corporation.

Hall, C. L., Guo, B., Valentine, A. Z., Groom, M. J., Daley, D., Sayal, K., & Hollis, C. (2020). The validity of the SNAP-IV in children displaying ADHD symptoms. *Assessment, 27*(6), 1258-1271.

Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck R. & Kerr, K. (2016). Implementation And Process Evaluation For Interventions In Education Settings: A Synthesis Of The Literature. Education Endowment Foundation. Retrieved 17 August 2021, from: https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting_up_an_Evaluation/IPE_Review_Final.pdf

Lane, C. et al. (2019). The cognitive profile of sotos syndrome. *Journal of Neuropsychology, 13*(2), 240-252.

Lendrum, A. & Humphrey, N. (2012). The importance of studying the implementation of interventions in school settings. *Oxford Review of Education*, *38*(5), 635-652.

Liew, J., Chen, Q., & Hughes, J. N. (2010). Child effortful control, teacher–student relationships, and achievement in academically at-risk children: Additive and interactive effects. *Early Childhood Research Quarterly, 25*, 51–64.

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental psychology,* 49(2), 270.

Nunes, T. et al. (2008). Improving children's working memory through guided rehearsal. *American Educational Research Association (AERA), New York: April*.

Nunes, T. et al. (2014). Improving deaf children's working memory through training. *International Journal of Speech and Language Pathology and Audiology*, 2(2), 51-66.

Nunes, T. et al. (n.d.). *Improving Working Memory and Arithmetic Knowledge*. Oxford: University of Oxford.

Nunes, T., Bryant, P., Evans, D., Bell, D., Gardner, S., Gardner, A. and Carraher, J. (2007). The contribution of logical reasoning to the learning of mathematics in primary school. *British Journal of Developmental Psychology, 25*, 147-166.

Peng, P., Namkung, J., Barnes, M., & Sun, C. (2016). A meta-analysis of mathematics and working memory: Moderating effects of working memory domain, type of mathematics skill, and sample characteristics. *Journal of Educational Psychology, 108*(4), 455–473.

Pickering, S.J. & S.E. Gathercole. (2001). *The Working Memory Test Battery for Children*. Hove: The Psychological Corporation.

Ricker, T. J., AuBuchon, A. M., & Cowan, N. (2010). Working memory. *Wiley interdisciplinary reviews: Cognitive science, 1*(4), 573-585.

Seitz, K. & Schumann-Hengsteler, R. (2000). Mental multiplication and working memory. *European Journal of Cognitive Psychology,* 12(4), 552-570.

Shipstead, Z. et al. (2010). Does working memory training generalize? *Psychologica Belgica,* 50(3), 245-276.

Singh, A. et al. (2019). *Understanding uncertainty in effect size calculation for multisite education trials*. Paper presented at EEF Analysis workshop, Durham.

Swanson, J. et al. (2001). Clinical Relevance of the Primary Findings of the MTA: Success Rates Based on the Severity of ADHD and ODD Symptoms at the End of Treatment. *Journal of the American Academy of Child & Adolescent Psychiatry*, *40*(2), 168-79.

Swanson, J. M. (1992) *School-Based Assessments and Intervention for ADD Students*. Irvine: KC Publishing.

Swinson, J. (2013). British Ability Scales 3. *Educational Psychology in Practice,* 29(4), 434-435.

Von Bastian, C. C., & Oberauer, K. (2014). Effects and mechanisms of working memory training: a review. *Psychological Research, 78*(6), 803-820.

Worth, J. et al. (2015) *Improving Literacy and Numeracy*. *Evaluation Report*. Education Endowment Foundation. Retrieved 17 August 2021, from: https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/Oxford_Numeracy_and_Literacy.pdf.

Wright, H. et al. (2019). *Improving Working Memory: Evaluation Report.* Education Endowment Foundation. Retrieved 17 August 2021, from: https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/Improving_Working_Memory_Report_final.pdf.

Zhai, F., Raver, C. C., Jones, S. M., Li-Grining, C. P., Pressler, E., & Gao, Q. (2010). Dosage effects on school readiness: evidence from a randomized classroom-based intervention. *The Social service review*, *84*(4), 615–655.

## Appendix 1: SNAP-IV Teacher Attention Rating Scale

**THE MTA SHORT FORM - ATTENTION RATING SCALE FOR TEACHERS**
Adapted from the original by James M. Swanson, University of California, Irvine

| Date | dd | mm | yy |
|---|---|---|---|

| | |
|---|---|
| Child's ID | |
| Child's name | |

| | |
|---|---|
| School | |
| School ID | |

**Instructions:** Think about the child's behaviour in the classroom during the last week. For each of the statements below, consider if the child's behaviour does **not at all** fit the description, or if it is **a little** like that, or is **mostly** like that, or if it is **very much** like that. Please put a black cross in the circle that best describes the child. Example: ☒

| | | Not at all | A little | Mostly | Very much |
|---|---|---|---|---|---|
| 1 | Fails to pay close attention to details or makes careless mistakes in schoolwork or tasks | ☐ | ☐ | ☐ | ☐ |
| 2 | Has difficulty sustaining attention in tasks or play activities | ☐ | ☐ | ☐ | ☐ |
| 3 | Does not seem to listen when spoken to directly | ☐ | ☐ | ☐ | ☐ |
| 4 | Does not follow through on instructions and fails to finish schoolwork or chores | ☐ | ☐ | ☐ | ☐ |
| 5 | Has difficulty organizing tasks and activities | ☐ | ☐ | ☐ | ☐ |
| 6 | Avoids, dislikes or reluctantly engages in tasks requiring sustained mental effort | ☐ | ☐ | ☐ | ☐ |

| | | | | | |
|---|---|---|---|---|---|
| 7 | Loses things necessary for activities (e.g. homework, books, toys, games, pencils, etc.) | ☐ | ☐ | ☐ | ☐ |
| 8 | Is distracted by extraneous stimuli | ☐ | ☐ | ☐ | ☐ |
| 9 | Is forgetful in daily activities | ☐ | ☐ | ☐ | ☐ |
| 10 | Fidgets with hands or feet or squirms in seat | ☐ | ☐ | ☐ | ☐ |
| 11 | Leaves seat in classroom or in other situations in which remaining seated is expected | ☐ | ☐ | ☐ | ☐ |
| 12 | Has difficulty playing or engaging in leisure activities quietly | ☐ | ☐ | ☐ | ☐ |
| 13 | Is always 'on the go' or often acts as if 'driven by a motor' | ☐ | ☐ | ☐ | ☐ |
| 14 | Talks excessively | ☐ | ☐ | ☐ | ☐ |
| 15 | Has difficulty waiting for their turn | ☐ | ☐ | ☐ | ☐ |

Name of teacher: ________________________________________

# Appendix 2: Privacy Notice

Privacy Notice

The delivery team (Oxford University) and evaluation team (RAND Europe, who will employ and supervise the work of a testing company, Alpha Plus, for the collection of outcome measures) conducting the IWM+A trial take privacy very seriously and, in light of the General Data Protection Regulations (GDPR), set out in this privacy notice how we will use and look after the personal information that we collect from the children and schools taking part in the trial.

Both the delivery and evaluation teams are the data controllers of the data they collect and are responsible for the processing of any personal data that is collected from those taking part in the trial within both teaching and comparison group settings. The two teams have put appropriate security measures in place to keep personal data secure and to prevent any unauthorised access to or use of it, including the use of password protected folders for digital information and locked filing cabinets for written records.

**What personal data will be collected, how, and what will it be used for?**

Personal data means any information about an individual from which that individual can be identified. Data for each eligible child in the study will be collected by both the delivery team and the evaluation team (with support from the testing company). The table below describes all the ways that children's personal data will be used by the delivery team and the evaluation team.

| **Personal Information Data** | | | | |
|---|---|---|---|---|
| **Data**<br><br>**(what will be collected?)** | **Source**<br><br>**(who will collect?)** | **Receiver or Permitted recipient** | **Purpose**<br><br>**(what will be used for?)** | **How do we collect the data?** |
| **Pupil data provided by the schools (data already held by schools):**<br><br>Administrative pupil data for Year 3 pupils participating in the study to be requested from schools:<br>- First name, surname, Unique Pupil Number (UPN), Free School Meal status (FSM), Date of birth, Gender | Oxford University (OU) (Data controller) will collect this from schools to allow for matching pupil data from the National Pupil database<br><br>(Data controller) | OU and RAND Europe<br><br>(Data controllers) | Matching pupil records; independent assessment of impact of intervention on outcomes; withdrawal forms; academic publications | Electronic data transfer |
| | | EEF<br><br>(Data controller)<br><br>and EEF archive manager (Data processor) | Long-term follow-up analyses and archiving of the data. EEF becomes data controller once the data has been transferred to the EEF Data Archive, but is not data controller until this has happened. | |
| **Pupil data from the National Pupil Database:**<br><br>- Free School Meal status (FSM) , Attainment data from the Early Years Foundation Stage Profile | National Pupil Database<br><br>(Data controller) | OU and RAND Europe (Data Controllers) | Independent assessment of impact of intervention on outcomes; longitudinal analysis of children's development; academic publications | Electronic data transfer |
| | | EEF<br><br>(Data controller) and EEF archive (Data processor) | Long-term follow-up analyses and archiving of the data. EEF becomes data controller once the data has been transferred to the EEF Data Archive, but is not data controller until this has happened. | |

| | | | | |
|---|---|---|---|---|
| **Pupil outcome data collected for the evaluation (collected by test administrators):**<br>- BAS3 test –subtest Number Skills<br>- GL PTM8<br>- Improving Working Memory Battery for Children – sub-tests Central Executive | RAND Europe<br>(Data controller)<br>(Data collection sub-contracted to Alpha Plus to test provider as Data Processor) | OU and RAND Europe<br>(Data controllers)<br>Alpha Plus<br>(Data processor) | Independent assessment of intervention impact on number skills and pupil's working memory, academic publications | Paper forms sent by secure mail; Electronic transfer via secure server |
| | | EEF<br>(Data controller)<br>and EEF archive (Data processor) | Long-term follow-up analyses. EEF becomes data controller once the data has been transferred to the EEF Data Archive, but is not data controller until this has happened. | |
| **Pre-test and outcome data collected for the evaluation (collected by schools):**<br>- Abbreviated version of SNPA-IV (Teacher rating scale of attention) | OU and Schools (Data processor)<br>OU and RAND Europe<br>(Data controller) | OU and RAND Europe<br>(Data controllers) | Independent assessment of impact of intervention on outcomes; longitudinal analysis of children's development; academic publications | Electronic transfer via secure server |
| | | EEF<br>(Data controller)<br>and EEF archive (Data processor) | Long-term follow-up analyses. EEF becomes data controller once the data has been transferred to the EEF Data Archive, but is not data controller until this has happened. | Electronic data transfer via secure server |
| **Data on delivery of IWM+A**<br>- IMWM+A games usage data (i.e. number of games played; number of games with 100% correct answers - collected by the delivery team through the game software)<br>- Highest level of game achieved with TA and online; paper data collected from schools<br>- Delivery team quality assurance records of TAs delivery of IWM+A sessions (children are present but no child data will be collected) | Schools (Data processor)<br>University of Oxford (Data controller) | OU and RAND Europe<br>(Data controllers) | Analyses of quality assurance and dose of intervention; academic publications | Electronic data transfer via secure server |
| | | EEF<br>(Data controller)<br>and EEF archive (Data processor) | Long-term follow-up analyses. EEF becomes data controller once the data has been transferred to the EEF Data Archive, but is not data controller until this has happened. | Electronic data transfer via secure server |

| | | | | |
|---|---|---|---|---|
| **Contact information**<br>- **Emails and names of TAs and link teachers** | Schools (Data processor)<br><br>University of Oxford (Data processor) | OU and RAND Europe (Data controller) | To maintain contact with schools and to send electronic questionnaires | Electronic data transfer via secure server |

**Uses of data obtained from NPD Linkage**

Further data will come from the National Pupil Database (NPD). The NPD is the government's official repository of data on all children and young people in state-funded schools. The evaluation team 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 and access to educational attainment and other relevant data. After the matching process has taken place, we will then analyse this data using the Secure Research Service (SRS) based at the Office of National Statistics (ONS). The SRS system does not allow users to remove or copy data from its servers.

**Who will hold the data?**

During the trial, the data will be managed by RAND Europe, Oxford University, and Alpha Plus. RAND Europe will securely transfer data to the EEF's archive at the end of the project under the EEF guidelines. At the end of the project, data will be submitted to the Office for National Statistics Secure Research Service (ONS SRS) for archiving in the EEF data archive (managed by FFT Education) and will include data only individually identifiable to the Department for Education.

Both the evaluation and delivery teams will seek ethical approval for the study. The study will not commence until ethical approval is received.

**How will the data be shared?**

During the trial data will be shared securely using a specialised software (Syncplicity). No data will be shared with processors outside the European Economic Area or the United Kingdom pursuant to EEA approved terms.

Pupil names will be shared with Alpha Plus to enable them to collect outcome data. For the purpose of research, following the completion of the trial, the data will be shared with the EEF archive who will act as the data controller after that point. You can read the EEF's Data Protection Statement in the EEF's website.[8]

[8] For more information about the EEF archive and Data Protection Statement, please see: https://educationendowmentfoundation.org.uk/public/files/Evaluation/Data_protection/Data_protection_statement_EEF_evaluations.pdf

**How will the data be kept secure?**

The delivery and the evaluation teams have put appropriate security measures in place to collect and to keep personal data secure and to prevent any unauthorised access to or use of it. All data will be collected and stored in accordance with the Data Protection Act (2018) and GDPR (General Data Protection and Regulation) requirements. No data will be saved on servers outside the European Economic Area or the United Kingdom pursuant to EEA approved terms. Evaluation data will be stored on secure servers. Data transferred between the delivery and evaluation teams will be encrypted or use secure file transfer protocols.

**What is the lawful basis we rely on in order to use and collect personal data and special category data?**

The data collected for the IWM+A project is used for the purposes of research. The lawful basis for OU and RAND Europe's use of that data under the GDPR is the 'legitimate interest' as the research will produce beneficial information for schools that will benefit children.[9] The delivery and evaluation teams will use data collected through the NPD linkage for the purposes of research. The evaluation team will use the basic identifiers (name and date of birth) to associate information with individuals to create datasets and that use will not affect the rights and freedoms of the children because this information will not be publicly connected to any other information. The test data will not be publicly available in association with the basic identifiers and therefore no risk of detriment exists to the children. Special category data is processed for the purpose of scientific research and archiving as permitted under GDPR Article 9 (j).

**How will existing data be archived?**

The evaluation team will archive data pertaining to the IWM+A trial as per the Table above into the EEF's archive in the Office for National Statistics Secure Research Service. All this data may potentially be linked to the NPD or to other data sources in the future for the purposes of research. When archived, this data has the potential to be used in further educational studies. No personal identifiers are submitted to the archive except for the Pupil Matching Reference, which is needed to link trial pupils with administrative data. The EEF archive has recently been transferred to the Office of National Statistics Secure Research Service, which significantly increases the security of the data. It can only be accessed by accredited researchers within the secure service. This fully complies with GDPR and the Data Protection Act 2018.

**How long will we keep data?**

Research data will be kept securely by the evaluation and by delivery team for the duration of the study and deleted one year after the report has been published. The delivery team will keep anonymised data (i.e. with no identification of pupils or schools and no date of birth, just age in months) that cannot be linked to any individual for three years after the publication of research papers, which is the standard period required for academic publications. Additionally, pupil data and data on the delivery of the intervention will be deposited into the EEF's data archive in the Office for National Statistics Secure Research Service at the end of all of the research.

[9] For more information about legitimate interest, please see: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/legitimate-interests/what-is-the-legitimate-interests-basis/

**What choices do you parents have in our use of their child's data?**

If parents do NOT want their child's data to be used, we will provide a withdrawal form for parents to return to their child's class teacher. During or after the intervention, participants may contact RAND Europe and/or University of Oxford to request deletion of their personal data.

**What are your rights?**

The evaluation and delivery teams will operate in accordance with EU law including GDPR. You are provided with certain rights that you may have the right to exercise through us. In summary those rights are:

- To access, correct or erase your personal data
- To object to the processing of your personal data

Were you to request information from us, we will need to confirm your identity to ensure the security of your data. We will endeavour to respond within 30 days, but our response time may vary depending on the complexity of your request.

If you wish to exercise any of these rights please contact the RAND Europe Data Protection Officer by emailing REdpo@randeurope.org or in writing to Data Protection Officer, RAND Europe, Westbrook Centre, Milton Road, Cambridge, CB4 1YG, UK. Once the project is completed and data has been shared with the EEF archive, participants should contact the EEF at info@eefoundation.org.uk

**Who can I contact about this project?**

Louise Matters – delivery team - wm.arithmetic@education.ox.ac.uk

Elena Rosa Brown – evaluation team – IWMplusA@randeurope.org