Education Endowment Foundation

IMPROVING WORKING MEMORY

Evaluation Report

May 2019

Hazel Wright, Richard Dorsett, Jake Anders, Jonathan Buzzeo, Johnny Runge and Michael Sanders

THE BEHAVIOURAL INSIGHTS TEAM

Education Endowment Foundation

The Education Endowment Foundation (EEF) is an independent grant-making charity dedicated to breaking the link between family income and educational achievement, ensuring that children from all backgrounds can fulfil their potential and make the most of their talents.

The EEF aims to raise the attainment of children facing disadvantage by:

- identifying promising educational innovations that address the needs of disadvantaged children in primary and secondary schools in England;
- evaluating these innovations to extend and secure the evidence on what works and can be made to work at scale; and
- encouraging schools, government, charities, and others to apply evidence and adopt innovations found to be effective.

The EEF was established in 2011 by the Sutton Trust as lead charity in partnership with Impetus (formerly Impetus Trust) and received a founding £125m grant from the Department for Education. Together, the EEF and Sutton Trust are the government-designated What Works Centre for improving education outcomes for school-aged children.

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

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

0207 802 1653

jonathan.kay@eefoundation.org.uk

www.educationendowmentfoundation.org.uk

Department for Education

What Works Network

The Sutton Trust

Impetus

Contents

About the evaluator

The evaluation team was led by Hazel Wright of the Behavioural Insights Team with Richard Dorsett of The University of Westminster and with support from Jake Anders of CREATE in the Department of Learning and Leadership, UCL Institute of Education. The process evaluation was conducted by Jonathan Buzzeo, formerly The National Institute of Economic and Social Research, (NIESR) and Johnny Runge (NIESR).

hazel.wright@bi.team

Executive summary

The project

This project tested the Improving Working Memory intervention (WM) and an adapted version, entitled the Working Memory Plus intervention (WM+). Working memory is the ability to remember and manipulate information over short time-frames. Previous research has suggested that working memory is a reliable predictor of numeracy outcomes.

The Improving Working Memory intervention aimed to improve the numeracy skills of Year 3 pupils (aged 7-8) who were behind the class average in numeracy by improving their working memory capacity. The intervention, developed and previously tested by a team at Oxford University, combined the explicit teaching of working memory strategies by Teaching Assistants (TAs) and the independent practice of these strategies using web-based games. The intervention was delivered in ten one-hour sessions and lasted for one term. The Working Memory Plus intervention also had ten sessions, but only five were focused on working memory, whilst the other five were focused on arithmetic content.

The project was a randomised controlled trial (RCT). 127 schools participated, being randomised at the school-level to one of three arms – the Improving Working Memory intervention, the Working Memory Plus intervention, or a business as usual control group. The primary outcome was maths attainment and the project also looked at working memory, and attention and behaviour in class as secondary outcomes. The process evaluation included fieldwork with eight intervention schools (four from each intervention), and an online survey of treatment and control schools. The trial took place between September 2016 and July 2017.

Key conclusions	
1.	Children in both the WM and WM+ schools made the equivalent of 3 additional months' progress in maths, on average, compared to children in the business as usual control schools. These results have high security ratings.
2.	Pupils eligible for Free School Meals (FSM) in the WM schools made a small amount of additional progress in maths compared to those in the control schools and no impact was found for FSM children in the WM+ schools. These results are of lower security than the overall findings because of the smaller number of pupils.
3.	The evaluation found positive impacts on working memory, and attention and behaviour in class for pupils receiving the interventions compared to children in comparison schools.
4.	The intervention was found to be time intensive, predominantly due to the need for TAs to leave class to deliver sessions, which increased pressure on teachers during lessons and in some cases required schools to source TA cover.

EEF security rating

This was an efficacy trial, which tested whether the intervention worked under developer-led conditions in a number of schools. The trial was a well-designed three-armed RCT. There were, however, some important differences in the pupils at baseline, with KS1 arithmetic scores higher for pupils in the control group and KS1 reasoning scores higher for pupils in the WM group. The trial security rating was therefore reduced to 4 padlocks.

Additional Findings

The primary outcome measures for both WM and WM+ are not statistically significant. This means that the statistical evidence does not meet the threshold set by the evaluator to conclude that the true impact was non-zero.[1]

Both interventions were found to have positive effects on working memory outcomes, measured using three tests - listening recall, counting recall and backward digit recall. Effects across all three were found to be larger in the WM group than the WM+ group. Positive effects were also found on attention and behaviour as measured by the Attention Rating Scale for Teachers.

Pupils reported that they enjoyed participating in both the WM and WM+ interventions. The process evaluation also looked at perceived impact. Some intervention schools suggested that there were improvements in pupils' abilities and performance (particularly in maths), their confidence, as well as their attitude and approach to learning in class.

Teachers reported challenges when delivering the interventions to pupils with pre-existing behavioural problems, with these pupils struggling to independently apply the strategies they were taught during the independent practice part of sessions and requiring more support. The intervention is not designed for pupils with special educational needs (SEN). However, several schools did include non-statemented SEN pupils in the groups selected to receive the interventions. The process evaluation highlights that the intervention is very time intensive. In particular, taking TAs away from class/pupils put additional pressure on classroom teachers, potentially with detrimental impacts on the rest of the class. Some Year 3 teachers sourced cover for the TA, which put additional demands on school resources.

Cost

The average cost was £23 per pupil per year when averaged over 3 years for the WM intervention, and £24 for the WM+ intervention. This estimate is based on delivery of the intervention to 12 pupils each year and includes direct costs, such as materials, equipment, travel and subsistence and photocopying. This estimate does not include costs associated with staff time such as TA cover, training and preparation, which are estimated at 52 days for a school working with 12 children receiving the WM intervention, and 54 days for those receiving the WM+ intervention.

Impact

Table 1: Summary of impact on primary outcome of maths (GL BAS3 Test)

Group	Effect size (95% CI)	Estimated months' progress	P value	No. of pupils	EEF security rating	EEF cost rating
WM vs. control	0.19 (-0.10, 0.47)	3	0.20	913	🔒🔒🔒🔒	£
WM+ vs. control	0.24 (-0.05, 0.52)	3	0.11	909	🔒🔒🔒🔒	£
WM FSM vs. control	0.09 (-0.14, 0.32)	1	0.49	281	n/a	£
WM+ FSM vs. control	-0.02 (-0.28, 0.23)	0	0.87	239	n/a	£

[1]The EEF uses a padlock rating system for trial security, which considers a number of factors that might limit the security of a finding. A full note on the EEF's position on statistical significance can be found here.

Introduction

1.1 Intervention

The Improving Working Memory project was a collaboration between the University of Oxford, the Education Endowment Foundation and the Behavioural Insights Team, and was designed to improve numeracy skills in year 3 pupils.

Current research within the educational literature has suggested that numeracy difficulties may be related to poor working memory capacity (Baddeley et al., 2011). On the basis of this body of research, the WM intervention was designed to help pupils who are behind the class average in numeracy by improving their working memory capacity. This intervention combined two strands of working memory training: strategies and practice. The first strand comprised a training programme taught by TAs in 10 structured sessions, which focused on strategies for improving working memory. The second strand focused on reinforcement of the strategies learned in the training programme through the use of computer games, which were designed specifically for the intervention.

A modified version of the intervention (referred to as the “Working Memory Plus” or WM+ group) that blended working memory activities with arithmetic content (from the mathematical reasoning curriculum) was also tested. In this version, five sessions with the TA were focused on working memory, and five were focused on arithmetic. The rationale for this was that pupils who are behind in their maths may require additional help with working memory, but also with maths-specific content. This blended intervention was delivered over the same time period.

Working Memory

In the WM arm, each child participated in ten one-hour sessions, alongside one other child. Each session consisted of 30 minutes of one-to-one TA-led activities and 30 minutes of working memory web-based games, played independently. The TA-led activities comprised a combination of three TA-led games, whilst three web-based games were used for the second half of the session.

During the TA-led training, TAs focused on teaching pupils’ strategies to remember. The strategies comprised two methods-oral rehearsal of the material the children needed to recall, and the use of spatial cues to support recall. For the first method, pupils were explicitly told by TAs that repeating a word they need to remember would help them remember the word more easily. Following this, while completing ‘Words’, ‘Colours’ and ‘Missing Digits’ games, pupils were instructed to orally rehearse the material that they needed to remember by repeating it aloud several times. The pupils were encouraged to rehearse orally until they remembered to do this independently. As the number of words to remember gradually increased, pupils were also taught to assign each word to a finger to help them recall the order of the words.

For the second method, the use of spatial cues, pupils were also encouraged to take note of where in space the target was and to point to it. This method was first taught using a ‘colours game’ in which pupils were tasked with remembering which colours appeared on a strip, and the order they appeared in. Pupils were taught to name the colour when it appeared and repeat this several times whilst pointing to the place where it appeared. As more colours were added to the strip, they were taught to rehearse the whole sequence this way from beginning to end.

A guided practice technique was employed for both of these methods, in which both TA and pupil would practice together until the pupil remembered to do this alone. To reinforce these methods, TAs also had children explain back to them the steps of the strategy, as this has been shown to encourage future use of a newly acquired technique.

The TA- led games: 'Words', 'Colours' and 'Missing Digits' were delivered via PowerPoint slides. All games were designed to train working memory. The 'Words' game was based on listening recall and contained a series of pictures each accompanied by a sentence. Pupils had to make a decision as to whether the sentence was true or false of the picture while also remembering the last word of the sentence. This task became progressively difficult as an extra sentence was added to each picture and the last word of each of these sentences had to be recalled. The 'Colours' game tested visuo-spatial recall by presenting pupils with a rainbow of colours in a specific order and then having them point to where each colour was positioned on a blank rainbow. Finally, the 'Missing Digits' game tested digit recall. Strings of numbers were presented on screen and pupils were asked to read aloud each digit and rehearse the last digit(s) of the string. This was done for multiple strings until a question mark randomly appeared. At this point, pupils were asked to recall the last digit(s) of the previous string.

Three web-based games pupils were asked to play independently during the other half of their session: 'Animals', 'Numbers' and 'Letters'. All of these games tested children's working memory capacity and required them to use the strategies that had been taught to them by TAs.

Game 1: "Animals", involved students counting target animals while ignoring non-target animals called Gremlins. At the end of each round pupils would have to recall the number of animals of each species (e.g. duck, monkey etc.) that they had counted. In game 2: 'Numbers', pupils were shown a series of number-filled grids with one number highlighted in each. After the presentation, pupils were given blank grids and asked to type in the numbers that had been highlighted in reverse order. Game 3: 'Letters', followed the same format as game 2, only this time pupils were asked to recall letters instead.

Each of the computer games had seven levels. In order to progress to the next level, the child had to successfully complete four trials. When a child failed to get four trials correct the TA moved onto a different game.

Working Memory Plus

In the WM+ arm, each child also participated in ten one-hour sessions. The first set of five sessions involved 30 minutes of TA-led working memory activities and 30 minutes of web-based games, played independently by the child (as previously described). However, sessions six to ten focused on number skills, comprising 30 minutes of TA-led number activities and 30 minutes of web-based number games that were played independently by the child.

The WM+ sessions focused on understanding additive composition of numbers and the inverse relations between operations, with the latter designed to promote the development of children's concepts of the operations of addition and subtraction.

The pupils played four TA-led additive composition games: 'Coins', 'Bags and Boxes', 'Gremlins' and 'The 7 and a half Game'. During 'Coins', the children were required to look at a picture of some coins and compose a series of amounts, with amounts becoming progressively more difficult. For instance, the pupil would be asked 'Imran needs to buy a stamp. What coins should he use to pay the exact money?'. Pupils were encouraged to start with the highest value coin, then to add to the next highest value until they reached the correct total. By doing so, pupils developed their understanding of equivalent values and improved their skills in counting on.

'Bags and Boxes' further developed their understanding of additive composition by extending the target amounts to numbers just under 1000. In this game, pupils were asked to compose a value by counting the value of boxes and bags on a screen (bags and boxes were each labelled with a value). As with the Coins game, pupils were taught to start with the highest value box or bag. During both games, once pupils had composed a total, they were at times also asked to compare two totals or match totals according to a further question.

'Gremlins' encouraged pupils to think of additive composition using positive and negative numbers. Pupils were shown that creatures called gremlins could be hit, as could spaceships. Pupils scored a point for each gremlin that was hit and lost a point for each spaceship that was hit. Pupils were taught to keep their score and encouraged to use different methods to calculate their score. The game comprised seven runs, which were played over more than one session where pupils did not complete all seven runs in the first session.

During the '7 and a Half' game, pupils were tasked with picking three cards from a shuffled deck and adding or subtracting the numbers to get as close as possible to a target of 7 and a half. By playing this game, pupils learned that there were many different ways of reaching the target. Pupils played against a fictional character called Jim, who drew his own cards and totalled his own score. They were asked to check his score and determine who won each round. As with 'Gremlins' pupils were asked to record their scores on a scoresheet.

The TA-led element of the session also included inversion games. These were designed to teach pupils about the inverse relation between addition and subtraction, and comprised six individual games: 'Blocks', 'Sequence Problems' 'More/less/the same', 'Just Numbers', 'Calculator Challenge' and 'Code Breaker'. In the first four games, pupils were presented with a series of scenarios, in which some items (e.g. ice lollies) were added or removed from a total that was either known or unknown. Where the original total was known, pupils were asked to calculate the final total, or determine what effect the changes would have to the original total, where it was unknown.

'Calculator Challenge' presented pupils with a series of scenarios and asked what calculation they would need to put into a calculator to solve the problem. While some problems were direct, inverse story problems were also presented, in which the original total is unknown (e.g. 'Tom has some keys in a wallet. We don't know how many. Mike gives Tom 4 keys. Now Tom has 12 keys. How many keys did Tom have before Mike gave him the new keys?'. Pupils were asked to use a calculator to show their working, with feedback presented on screen to show them the correct solution after each game.

During 'Code Breaker', pupils were given opportunities to practice the inverse relation between operations using a calculator, to determine an original value that is unknown. For instance, pupils were asked "The spy thought of a number, then he added 35 and the result was 86; what number was he thinking of?". As with all games, the pupils received feedback demonstrating the correct solution after each attempt.

The remaining 30 minutes of the session was always spent playing computer games designed to allow the pupils to practise the concepts learned during the TA session. While these were played independently, pupils were guided by the TA. Three or four games were timetabled each week, and included a mixture of additive composition games and inverse relation games. Unlike the games played in the WM group, pupils did not need to pass a certain number of questions to progress to the next game.

Intervention Delivery

Intervention sessions were always divided in two halves as TAs took children out of class in pairs. TAs were also provided with a script to ensure that the training was uniformly delivered. For the first half of the one-hour session, one child played the computer games independently while the other child worked with the TA. This was reversed for the second half of the session.

Progress through the independent game element of each intervention was based on each child's performance. While all children started the intervention at the same point, they could only move through levels of the game once they had reached the criterion for the previous level. In the WM+ arm, all the children also started the arithmetic games at the same level. This means that children who progressed faster had the opportunity to play more games than those who progressed at a slower pace.

All sessions took place at a dedicated period during the school day which was additional to maths teaching. In the event that a session was missed, an extra catch-up session was provided where possible.

For both the WM and WM+ group, TAs were specially trained prior to implementation and a link teacher belonging to each school was nominated to monitor and support implementation. All TAs and link teachers participated in a one-day preparation session which was delivered by the Oxford project team. During this session, the intervention materials were introduced and attendees were shown how to set up the games and teach the working memory strategies. Those in the WM+ group received one day of extra training, which mirrored the approach used for the WM training. This included an introduction to the concepts of arithmetic they would be teaching, a presentation of the games used to deliver the programme and opportunities to practice playing and assessing the games.

In addition to their training, all TAs were provided with a handbook, designed by Oxford University, that contained detailed delivery instructions for both the TA-led activities and the independently played computer games. The delivery team provided ongoing support and advice to schools throughout the programme, with visits scheduled half way through the programme to allow the delivery team to provide extra advice and feedback to TAs.

In a slight change to the original protocol, during recruitment it was determined that English state primary schools with at least 18 pupils in year 3 would be eligible for the trial, instead of the 20 originally indicated in the protocol. These schools were recruited by the University of Oxford. The intervention took place on school grounds. A quiet teaching area where children could listen without being distracted and could speak aloud without disrupting others was set-up. This workspace contained two computers with internet access on separate tables to allow one child to follow the strategy training and a second child to independently practice the computer games. Although children attended the sessions in pairs, intervention delivery was on a one-to-one basis.

The intervention was not modified during the course of the study. However, it should be noted that the selection criteria for pupils were updated early in the intervention to make clear that SEN pupils should only be excluded if they had been statemented.

As part of the process evaluation, the intervention adherence and fidelity were assessed through findings from the fieldwork visits to eight treatment schools and the end-of-project survey. Generally, the qualitative findings demonstrate a high level of fidelity. It suggests that treatment schools did not make any significant adaptations to the intervention, with TAs in most cases adhering strictly to the handbook instructions. Regarding the length and structure of sessions, the intended dosage for the trial seems to have been largely met, with schools generally delivering the required number of weekly sessions and maintaining a 50/50 split between TA-led and online games.
The process evaluation also examined issues relating to fidelity and implementation. The qualitative findings show that treatment schools did not make any significant adaptations to the intervention, with TAs in most cases adhering strictly to the handbook instructions. Regarding the length and structure of sessions, the intended dosage for the trial seems to have been largely met, with schools generally delivering the required number of weekly sessions and maintaining a 50/50 split between TA-led and online games.

However, a few factors in the school environment were seen to affect how the intervention was delivered. Firstly, most treatment schools had timetabled either exclusively or mostly in the afternoon. Some interviewed TAs found that this did not always get the best out of pupils who tend to concentrate better in the morning, but core subject teaching was prioritised and some schools did not have access to the necessary resources in the mornings in any case. Second, schools reported mixed experiences in securing the necessary resources, namely two working computers, internet access and a quiet teaching area. The latter affected pupils' ability to complete independent online games and schools that had to search for suitable spaces on an ad hoc basis often had shortened sessions. Another factor that affected implementation was how easy/difficult it was for pupils to work independently through the online games for 30 mins. In this regard, TAs reported varying experiences. When there were problems, TAs attributed this to behavioural issues and lack of IT skills. It was also noted that in five schools the trained TA became ill and had to be replaced either to the end of the project or temporarily. In these cases, new TAs were also provided with training.

Finally, the process evaluation identified some fidelity issues regarding the selection of pupils. Most schools did follow the instructions from the delivery team. The majority of schools chose between 10 and 20 pupils to participate (most chose the minimum due to the burden on staff time placed by the intervention), as well as selecting those with lower number skills and excluding those with statement of SEN. However, some treatment schools reported having at least one SEN pupil receiving the intervention, though the majority of these were not statemented. In any case, in some instances, difficulties were reported relating to SEN pupils receiving the intervention, in particular their ability to maintain concentration and work independently through online games.

1.2 Background evidence

Working memory is the ability to remember and manipulate information over short time-frames. Training to improve memory as a means of increasing attainment has foundations in cognitive science (Baddeley, 2000). Working memory has been shown to be a reliable predictor of attainment in numeracy (Baddeley et al., 2011).

Early evidence for the positive effects of computer-based working memory training was provided by Holmes et al. (2009) in a small sample[2] of ten-year-old children with learning difficulties. Using the adaptive[3] version of the Cogmed training programme, the authors reported significant improvements in all aspects of working memory tested including: verbal short term memory ($p = .01$, $d = .62$), visuospatial short term memory ($p < .01$, $d = 1.20$), verbal working memory ($p < .01$, $d = 1.55$), and visuospatial working memory ($p < .01$, $d = 1.03$) for children completing this form of training. Improvements in three of these aspects of working memory also endured six-months post training: visuospatial short-term memory ($MSE = 112.93$, $p < .01$), verbal working memory ($MSE = 110.16$, $p < .01$), and visuospatial working memory ($MSE = 121.74$, $p < .01$). In addition, compared with pre-training baseline levels, significant improvements in mathematical reasoning scores ($F(1, 17) = 9.50$, $MSE = 48.66$, $p < .01$) emerged for children who had received the adaptive training.[4] This indicates that computer-based working memory training may be effective in helping children to surmount working memory impairments and associated learning difficulties.

Later, Holmes and Gathercole (2014) tested the use of Cogmed in a classroom setting using a larger sample of 50 children with low academic performance. Pupils completed a minimum of 20 group training sessions with their teacher. These pupils were matched against children who had received no training. The authors reported training as being associated with significant improvements in maths scores for children in years 5 ($d= 1.15$) and 6 ($d= 0.60$) respectively. Significant improvements were also observed in English scores for year 6 ($d= 0.67$)[5]. These findings provide evidence that working memory training can provide benefits under real conditions, such as the classroom, with relatively large groups of children.

More recently, Passolunghi and Costa (2016), trialled a programme of working memory to develop early

[2] There were 22 children in the group which received the adaptive training and 20 in the ‘comparator group’ who received non-adaptive training.
[3] The adaptive version of the program matches task difficulty to the child’s current memory span on a trial-by-trial basis for each task. With the non-adaptive version, difficulty levels are fixed at a low-level for the duration of training.
[4] This improvement in mathematical reasoning only emerged in the 6-month post-training.
[5] Significance levels not clearly reported in paper.

numeracy skills in 5-year-old preschool children. The authors compared the effect of general working memory training against specific numeracy training. The working memory sessions comprised a combination of interactive games training visuospatial, verbal and episodic memory, while the numeracy training targeted specific mathematical skills such as counting and quantity matching. After five weeks of training, only those in the working memory group demonstrated working memory enhancements both in terms of visuospatial working memory (Mdiff = 2.52, p < .001, d = 1.16), and verbal working memory (Mdiff = 1.96, p = .002, d = .97). However, both groups exhibited significant improvements in early numeracy abilities (working memory group: Mdiff = 3.82, p = .005, d = .80; numeracy training group: Mdiff = 6.65, p < .001, d = 1.63) relative to the control group. These findings indicate the importance of cultivating domain general abilities, such as working memory, alongside more specific abilities, like numeracy skills, to help prevent future learning difficulties. These findings run parallel to previous work by Alloway (2009) which highlighted poor working memory ability as being tightly related to poor performance in maths.

The WM programme used in this study has been developed and tested in previous studies by Oxford University. One of the first of these studies tested the use of guided rehearsal in improving working memory - in particular, the central executive component (WM-CE) which is required in a variety of school tasks (Nunes, Evans, Bell & Campos, 2008). Thirty-five normally developing children aged between five and seven years took part in the study and were randomly allocated to either a WM-CE training or a control group. The training comprised three half-hour sessions delivered with the support of a computer and was designed to promote different aspects of metacognitive skills including listening recall, counting recall and backward digit recall training. The results revealed a significant effect of the training on all three WM-CE outcome measures. Cohen’s d effect size was 1.2 SD for counting recall, 1.5 SD for listening recall, and 1.2 SD for backward digit recall--all very large effect sizes. Another study using a similar technique with deaf children found significant improvements in working memory scores as a result of training (d=0.78, p<.001) (Nunes, Barros, Evans & Burman, 2014). Although both studies had small sample sizes of 35 children respectively, they provide promising results.

The WM+ arm of the intervention includes elements of the Mathematics and Reasoning Programme, also developed by the Oxford University team. These elements aim to develop children’s understanding of the logical principles underlying mathematics. In a previous EEF evaluation with pupils in year 2 (Worth et al., 2015) this programme was found to have a positive impact on pupils’ numeracy ability. This was equivalent to three additional months progress over the course of a year, with an effect size of 0.2 standard deviations on mathematics achievement.

Given these previous findings, the current intervention was deemed ready for an efficacy trial with attainment and working memory tested as outcomes.

1.3 Evaluation objectives

The impact evaluation sought to answer the following research questions:

- What is the effect of Working Memory and Working Memory Plus on children's number skills at the end of year 3, as assessed by a subset of questions from the British Ability Scales 3rd Edition?
- What is the effect of Working Memory and Working Memory Plus on participants' working memory at the end of year 3? This was assessed at the end of year 3 using the three central executive subtests (counting recall, backward digit recall, listening recall) of a working memory scale for children (Pickering & Gathercole, 2001, or Alloway, 2007, which is the computerised version).
- What is the effect of Working Memory and Working Memory Plus on participants' attention and behaviour in class at the end of year 3? This was assessed by teachers, who were asked to complete 15 items of the "Attention Rating Scale for Teachers" (adapted from the original by James M. Swanson; Swanson et al., 2001).

The effectiveness of the intervention for pupils eligible for free school meals (FSM) was also explored, as standard in all EEF trials.

The purpose of the process evaluation was to assess the intervention in terms of fidelity to the programme intentions, and the scalability of the programme. The delivery and evaluation of this project was funded as part of the EEF's routine round of grants.

The Evaluation Protocol can be found here.

1.4 Project team

The project team comprised Terezinha Nunes, Peter Bryant, Rossana Barros Baertl, Deborah Evans and Susan Baker of Oxford University. The intervention was delivered by teaching assistants trained by The University of Oxford team. The intervention was designed by Terezinha Nunes, Deborah Evans, Rossana Barros and Peter Bryant.

The evaluation team was led by Hazel Northcott of the Behavioural Insights Team with Richard Dorsett of The University of Westminster and with support from Jake Anders of CREATE in the Department of Learning and Leadership, UCL Institute of Education.

Jonathan Buzzeo, formerly The National Institute of Economic and Social Research, (NIESR) and Johnny Runge (NIESR) carried out the process evaluation.

1.5 Ethical review

The project was approved in January 2016 by the University of Oxford Ethics Committee. As randomisation took place at school rather than individual level, consent from the school was sought for this. A parental information sheet, provided with a consent form, gave information on the aims of the research and the use of data to allow parents to make an informed decision regarding consent for data sharing. The opportunity of opting-out of disclosure of their Unique Pupil Identifier was also part of the consent process.

1.6 Trial registration

This trial has been registered at: http://www.controlled-trials.com/ISRCTN47105456.

Methods

2.1 Trial design

The evaluation was run as a randomised controlled trial with 1475 pupils in year 3 (aged 7-8 years) across 127 primary schools, randomised at school level to three groups - the Working Memory Intervention (42 schools), the Working Memory Plus (41 schools) intervention and a control condition (44 schools). The control condition comprised a ‘business as usual’ approach where schools continued with normal classroom teaching and support for eligible pupils.

This design was selected to maximise statistical power, while reducing the potential for contamination which would have been more likely had pupil- or class-level randomisation been used.

Pupils in the intervention groups were intended to receive ten one-hour sessions over one school term; pupils in the control group experienced their usual teaching. The programme was provided as additional to pupils’ usual maths lessons, displacing time usually spent in the afternoon in other lessons such as PE, history, science, art, and geography.

2.2 Participant selection

English state primary schools were eligible to participate in the trial, provided they had at least 18 pupils in year 3. Schools were recruited by the delivery team at the University of Oxford and were located across England, with small clusters in Oxford, Sheffield and Leeds. There were several methods employed to recruit schools, with some directly approached through letters inviting them to participate and others indirectly approached through advocates. Invitations were also published on the website of the Department for Education, The University of Oxford, and the website of the Oxford University Press. Finally, fliers about the project were distributed at different events in which members of the team had the opportunity to meet school representatives (e.g. teachers, numeracy experts).

Teachers were asked to nominate year 3 children prior to randomisation, who were in the lower third or lower half of their class, by their KS1 arithmetic attainment.
Depending on the number of pupils that were in year 3, they were asked to select between 10 and 20 pupils to participate. Schools with fewer than 20 pupils in year 3 were the exception, and were allowed to nominate a minimum of 9 to participate. At a later point, and before randomisation, the decision was also taken to exclude SEN statemented children. This was due to potential complications associated with their engagement in the intervention and with test procedures.

Schools wishing to participate in the trial were asked to sign a Memorandum of Understanding (Appendix D) which committed them to full compliance with the requirements of the trial, including supplying the necessary pupil data. Opt-out consent was sought from parents of all eligible pupils for agreement for data sharing, also prior to randomisation. Schools in the control group were given a financial incentive of £725 to be spent on the training, if desired, after the post-test.

2.3 Outcome measures

The intervention aimed to help pupils who were performing below the class average in numeracy and was tailored to support pupils to do this by improving their working memory. The primary outcome was pupils’ numerical skills, as assessed by the standardised GL Assessment British Ability Scales 3rd Edition numeracy skills test. For the purpose of this trial, questions 7-18 were used, as these are appropriate for the age range 7-10 years inclusive. The tests were administered by RAs at the end of year 3. The RAs were recruited and trained by BIT and blinded to allocation status. The tests were conducted individually, with the test subject sat opposite the RA and the test conducted under exam conditions. No members of the school staff were involved in test administration, and every effort was made to ensure RAs were blind to treatment allocation. RAs were also responsible for marking each question as the test was administered, ensuring that marking was also blind to group allocation. Total scores were calculated by summing the number of questions answered correctly by pupils. As per the Evaluation Protocol, pre-test scores were not collected, with KS1 arithmetic and reasoning attainment (assessed during year 2) used as baseline measures instead. Final analysis of our primary outcome (BAS3 scores) included all pupils for whom we had collected BAS3 scores, with multiple imputation employed for pupils for who were missing either both, or one of the KS1 baseline scores. Final analysis included 1352 pupils.

The secondary outcomes of interest in this trial were pupils' working memory and their attention and behaviour in class. Working memory was assessed prior to randomisation and hence before the intervention was administered, and at the end of year 3 using the three central executive subtests (counting recall, backward digit recall, listening recall) of a working memory scale for children, a standardised measure validated for the UK population (Pickering & Gathercole, 2001, or Alloway, 2007). In each test, two practice trials were delivered to ensure pupils understood the approach, followed by four trials at each difficulty level, with difficulty increasing at each level. For each test, the total number of correct trials was summed to create a score.

The Counting Recall subtest required pupils to count dots on subsequent pages of a book, and later recall the number of dots on each page. During the Listening Recall subtest, pupils heard a sentence, had to judge whether it was true or false, and were asked to try to remember the last word of each sentence. In the Backward Digit Span test, the pupils listened to a series of digits and were asked to recall them in the opposite order.

Each subtest was administered by RAs trained by Oxford University, blinded to allocation status. RAs from BIT audited these assessments at three schools, each for one full day of testing (between 18 - 25 tests) to ensure that they were completed as per the protocol and that assessors were blind to the allocation of the child. Working memory scores were also collected at different points. Children in the WM group participated in the programme for 10 weeks and were tested after its completion. Children participating in the WM+ group participated in the working memory training for five weeks and were tested after they had completed the number training, which extended for five weeks after the Working Memory training was completed. Thus, the assessment of this group was a delayed post-test, five weeks after the working memory portion of the programme has been completed.

Attention and behaviour in class was assessed by teachers both at pre-test and at the end of year 3. When testers visited schools for data collection on the working memory test, they approached teachers, who were asked to complete 15 items for the "Attention Rating Scale for Teachers" (adapted from the original by James M. Swanson; Swanson et al., 2001). This is a 4-point rating scale which contains items relevant to children's sustained attention in the classroom. Given teachers were assigned responsibility for completing this survey, and that they could not be blinded to the treatment allocation of their school, this measure introduces some threat to internal validity.

2.4 Sample size

The project team aimed to recruit 115 schools. Within schools, it was expected that there would be an average of 15 eligible year 3 pupils, so approximately 1,700 pupils in total. Power calculations were based on a simplifying assumption of 16 eligible pupils per school. Further assuming 88% would be observed in the data (this is informed by Worth et al., 2015) reduces this to 14 useable pupils per school, on average. In light of the Worth et al., (2015) results, we assumed an intra-cluster correlation of 0.12 and that the pre-test would account for 57% of outcome variation. The minimum detectable effect size (MDES) for a 2-tailed test with 95% significance and 80% power was estimated under these assumptions to be 0.18 for a 3-arm trial with a third of schools in each arm.[6]

[6] Our assumptions are summarised in Table 9 of this report.

In fact, the project team exceeded the target, recruiting 127 schools. Despite this, the number of pupils involved in the trial was less than expected. This reflects the uncertainty in the original estimate of numbers eligible within each school (there were fewer than 12 on average, rather than the assumed 16) but is unlikely to have a noticeable effect on the power of the analysis since this derives primarily from the number of schools recruited.

2.5 Randomisation

Schools recruited by the project team were randomly assigned by the evaluation team. Among the 127 schools, there were two pairs of schools that had to have the same treatment allocation. Two of these schools were federated and thus paired to form one unit for randomisation. Two further schools shared the same SEN Coordinator (SENCO), and thus were also paired to form one unit for randomisation. By treating each such pair as if it were a single school for randomisation purposes, we effectively reduced the number of units to 125. Schools were then categorised on the basis of their size (one-form entry, two or more form entry; weighted averages were used in the case of the combined school-pairs) and most recent school KS1 performance, defined as the % of all year 2 pupils assessed as working at the expected standard or at greater depth in maths in Summer 2016 (lower third, middle third and upper third). This resulted in six blocks and schools were randomised within each block. The purpose of this blocking was to improve cross-arm comparability of schools and also to increase precision of estimates. Since there were more two-form entry schools than one-form entry schools, the block sizes for the former were larger than those for the latter (27, 26, 26 schools across the upper, middle and lower thirds respectively compared to 15, 17 and 14 schools.)

The randomisation was carried out using the statistical package Stata. Randomisation was implemented in a way that achieves an equal number in each arm:

- Each school was assigned a randomly generated number
- Schools were sorted by block and random number
- The first school was assigned randomly to one of the arms
- Successive schools were assigned in the following order: WM arm, WM+ arm, control arm.

This continued until all schools had been assigned.

The computer code used to carry out the randomisation is reported in Appendix C.

2.6 Analysis

Impacts were estimated using linear regression models. Outcomes were regressed on dummy variables indicating a whether a school was in the WM or the WM+ arm. The estimated coefficient in each case captures the impact of a school being in the respective treatment arm rather than in the control group. Block indicators were also included in these models, along with the appropriate pre-test(s). In the case of the primary outcome, BAS3 score, there was no equivalent pre-test but instead two baseline measures were used: year 2 arithmetic and year 2 reasoning scores. For the secondary outcomes, the three working memory subtests were administered pre-randomisation by researchers from the University of Oxford, whilst the pre-tests for attention and behaviour in class were delivered by teachers.

As this is a school-level randomised controlled trial, inference was based on standard errors adjusted for school-level clustering using Stata's 'cluster' option. This is in acknowledgement of the likelihood of outcomes being correlated within schools. Clustering standard errors in this way is reasonable given the large number of schools involved. Furthermore, it is an attractive approach relative to the leading alternative of a multilevel model since it avoids the need to assume that school-level effects are uncorrelated with other regressors and the biases that can result when this assumption is not met (Ebbes et al., 2004).

The level of missingness among the KS1 regressors in the full sample was 6%. The evaluation protocol did not specify how missingness would be handled. We have used multiple imputation (as implemented by Stata, with 10 imputations) to allow estimation to proceed on the basis of all pupils for whom we have outcome data. Multiple imputation assumes missingness is at random conditional on observable covariates, something that we cannot verify. Imputation in this case is on the basis of treatment arm and blocking dummies. Despite this being unlikely to control for the influences generating missingness, multiply imputing serves to allow the resulting estimates to reflect the additional uncertainty resulting from some of the baseline measures being unknown. Furthermore, it avoids the loss of sample that proceeding without imputation would have entailed, or the loss of power associated with not using the baseline measure. As sensitivity checks, we also used single null imputation which gave qualitatively similar results albeit appearing more precise; and an analysis only with complete cases. This precision is illusory, stemming from not taking adequate account of the missing values. We note that this also assumes missingness to be at random. Appendix E includes those results as well as the complete-case results.

The regression results capture the effect of intention to treat. Estimates are presented as effect sizes, calculated using the Hedges' g formula, expressing the estimated effect as represented by the regression coefficient relative to the total unadjusted outcome variance in the sample.

Formally, the effect sizes are calculated as follows

$$g^{*} = \frac{\Gamma((n_T + n_C - 2)/2)}{\sqrt{(n_T + n_C - 2)/2} \cdot \Gamma((n_T + n_C - 3)/2)} \cdot \frac{\beta_T}{\sqrt{\frac{(n_T - 1)s_T^2 + (n_C - 1)s_C^2}{n_T + n_C - 2}}}$$

where n_T is the number of treatment group observations (WM or WM+, as appropriate), n_c is the number of control group observations, $\Gamma()$ is the gamma function, β_T is the regression coefficient on the dummy variable indicating membership of the treatment group, S_T^2 is the standard deviation of the outcome variable among the treated group and S_C^2 is the standard deviation of the outcome variable among the control group.

Impacts were also estimated for the subgroup of FSM pupils following the same model specified above. This subgroup was identified using the NPD variables everfsm_6_p_spr17.

Games Data Analysis

In order to understand the implementation of the intervention better, we conducted exploratory analysis on the "Games Data", which is audit data produced by the Improving Working Memory online software, logging which games individuals play, when they play them, and their score on the game. As we have pupil and school identifiers we are able to describe how many games individuals play and how this varies between schools. For obvious reasons, these data are only available for individuals in the two treatment groups. As such, it should be remembered that we are making no causal claims from these but rather present them as part of understanding the implementation of the intervention.

As such, we conducted primarily descriptive analysis on these data. In addition, this analysis was not pre-specified in the evaluation protocol, so it should be treated as exploratory only. The analysis consisted of:

- Descriptive statistics, histograms and kernel density plots by treatment arm of the number of games individuals play over the course of the treatment period
- Descriptive statistics, histograms and kernel density plots by treatment arm of the number of games played in each school over the course of the treatment period
- Linear regression analysis of the number of games played between the two treatment arms estimated on person-level data:

- $NumGames_i = \alpha + \beta_1 Treat_i + \varepsilon_i$
- where $NumGames_i$ is the number of games an individual (i) plays during the treatment period, $Treat_i$ is a binary indicator of treatment arm (WM+ vs. WM) and ε_i is an idiosyncratic error term with standard errors calculated taking into account clustering at the school level.
- As such, β_1 recovers the difference in mean number of games played associated with being in the WM+ rather than the WM group.

● Linear regression analysis of the number of games played between the two treatment arms estimated on school-level data:

- $NumGames_j = \alpha + \beta_1 Treat_j + \varepsilon_j$
- where $NumGames_j$ is the number of games played in a school (j) during the treatment period, $Treat_j$ is a binary indicator of treatment arm (WM+ vs. WM) and ε_j is an idiosyncratic error term with standard errors calculated taking into account clustering at the school level.
- As such, β_1 recovers the difference in mean number of games played in a school associated with being in the WM+ rather than the WM group.

● Linear regression analysis of game score on how many games an individual has played by this point using the three models estimated on game-level data as follows:

- $y_{ij} = \alpha + \beta_1 GameNum_{ij} + \varepsilon_{ij}$ to identify raw change in performance as an individual plays more games (β_1)
- $y_{ij} = \alpha + \beta_2 GameNum_{ij} + \eta_j + \varepsilon_{ij}$ to identify change in performance as an individual plays more games of the same type (β_2)
- $y_{ij} = \alpha + \beta_3 GameNum_{ij} + \varsigma_j + \varepsilon_{ij}$ to identify change in performance as an individual plays the same game multiple times (β_3)
- where y_{ij} is in-game performance by person i in game j, $GameNum_{ij}$ indexes the number of games the individual has played, η_j is a vector of game type identifiers, ς_i is a vector of game-specific identifiers, and ε_{ij} is an idiosyncratic error term with standard errors calculated taking into account clustering at the school-level.

● Descriptive statistics and cross-tabulation by treatment arm of the number of TA sessions and Computer Games sessions in which it is reported pupils participate.

2.7 Implementation and process evaluation

The overarching purpose of the process evaluation is to show how the IWM intervention was implemented by treatment schools, whether this differed from the intended treatment model, and the factors that informed this. It also monitors the activity of the control group to establish what is being done in the absence of the intervention. In doing so, the process evaluation aims to bring greater clarity to the quantitative research findings and to understand the reasons behind the impact findings.

Specifically, this process evaluation sets out to understand the following aspects of implementation:

- Implementers' experience of the training they received and their level of preparedness for delivering the intervention
- Whether participants (pupils) were selected appropriately by treatment schools (fidelity), and how much of the intended intervention was delivered (dosage)
- How responsive pupils were to the intervention
- What adaptations (if any) were made by implementers
- How the implementation environment affected the quality of delivery

The process evaluation also seeks to highlight the perceived impact of the IWM intervention in the eyes of implementers/participants, and to gather their views on how the intervention can by improved to inform its future rollout.

The following research methods were used to understand these aspects:

- Fieldwork visits to eight treatment schools (four in each treatment group), conducted from March to July 2017. Fieldwork visits included:
  - Interviews with eight link teachers
  - Interviews with seven teaching assistants (TAs)
  - Observations of six one-hour intervention sessions

In addition, data was collected via two surveys:

- An end of project survey, administered from June to July 2017
- A control group survey, administered from June to July 2017

The end of project survey was completed by at least one staff member from 74 per cent of all participating schools. In some cases, we received several responses from each school. In total, 127 participants completed the survey, consisting primarily of link teachers and TAs. The control group survey was completed by just over half of schools allocated to the control group (24 out of 43). Respondents mainly included year 3 teachers, SENCOs and maths leaders.

Fidelity and dosage was primarily explored in the survey. This included questions on the number and type of pupils selected to participate, and the length and number of sessions per week. In addition, the nature of any adaptations and the implementation environment were explored in both the survey and during the fieldwork interviews.

It should be noted that the views and experiences presented are not necessarily representative of all treatment and control schools. Surveys were not responded to by all schools and in some instances by staff members with different roles within the school and the project. Further, while steps were taken to ensure that the treatment schools visited as part of fieldwork included a variety of delivery contexts, the number that participated in this part of the research was very small compared to the larger treatment population. As such, some perspectives on the delivery of the improving working memory project may be missing from the final analysis.

All activities were carried out by NIESR with support from BIT and the delivery team at the University of Oxford.

2.8 Costs

Cost information was collected from a number of sources. Primarily, costs were estimated by the lead independent evaluator. The project team provided additional information on costs related to TAs and equipment. The Qualitative evaluation team also contributed insights collected during their evaluation of the intervention via interviews and observation.

RAs trained and employed by BIT collected survey data from schools directly, using a survey designed by the independent evaluators for this purpose. The independent evaluators specified direct delivery costs as those relating to materials, travel and subsistence, and printing. Costs related to staff time, staff cover and training were recorded separately. Costs related to staff time for the coordination and implementation of the trial were not included in the final estimate.

The cost survey data was collected face to face from teachers, following BAS3 outcome collection by RAs. RAs sat with teachers to discuss the items on the survey and recorded the results on paper. The cost survey was estimated by RAs to take approximately 20 minutes to discuss and deliver.

This data was collated by the lead evaluator and cost calculations were carried out, assuming that a school requires two teachers or TAs to deliver the intervention. Estimates were calculated on the basis of teachers/TAs working with 12 pupils over one academic year, given the average number of pupils participating in the trial per school was calculated at 11.6. Costs were estimated separately for schools in the WM and WM+ arm. In each case, start-up costs per school, running costs and running costs per pupil were specified.

Staff costs were calculated in hours using estimates of the number of hours required for teachers to attend training, for staff cover and for intervention delivery. Training and delivery estimates were based on information from the Project Team and qualitative evaluators and later corroborated by information collected via the face to face surveys with schools, who also provided information on time spent preparing for the delivery of the intervention.

2.9 Timeline

Table 2: Timeline

Date	Activity
Jan - August 2016	Development of WM+ intervention by The University of Oxford Recruitment of schools by The University of Oxford Ethical approval obtained by The University of Oxford
Sept - October 2016	Identification of pupils by teachers Working memory pre-tests administered by The University of Oxford Opt-out parental consent obtained by The University of Oxford
Nov - Dec 2016	TAs trained by The University of Oxford
Dec 2016	Linkage to KS1 pre-test by Behavioural Insights
Jan - May 2017	Programme delivered
Oct 2016 - May 2017	Process evaluation delivered by NIESR
May-July - 2017	Numeracy post-tests delivered by Behavioural Insights Cost data collected by Behavioural Insights Working memory and attention and memory post-test administered by The University of Oxford
Aug - Sep 2017	Analysis
Dec 2017 - March 2018	Reporting

Impact evaluation

3.1 Participants

The project team initially approached 387 schools to determine whether they would be interested in participating in the trial. This initial approach was made via a speculative email, or a letter, sent from the Oxford University team directly to eligible schools. A form was included for schools to complete and return, confirming that they would like to participate. Participant flow for our primary outcome measure is detailed in Figure 1 below.

The minimum number of children in the year group had to be at least 18 per school for the school to take part in the project. 173 completed forms were returned to the project team, and of these, 9 schools were not eligible due their size. The remaining 164 schools that were interested were sent a Memorandum of Understanding (MOU) to complete. 144 MOUs were received, and the project team started a waiting list for schools who responded after the desired total number of schools for the trial had been achieved.

16 schools changed their minds for various reasons, primarily due to staffing changes that happened after they had sent the interest form. This was mainly due to key members of the teaching and non-teaching staff leaving the school or in the case of one school, due to this school going into special measures. Spaces created by these schools were filled with schools who had been placed on the waiting list. This left 128 schools who had signed MOUs, with one further school dropping out prior to pre-test due to staffing changes, leaving 127 schools to participate in the trial. School characteristics are set out in tables 3-6 below.

Details of the school characteristics show some substantive differences between groups of schools allocated to each arm. The WM group comprised nearly twice as many schools with ‘Outstanding’ Ofsted ratings as the other two groups (table 3), though this data was unobtainable for a small proportion of the control group schools.

On average, a much higher proportion of the schools allocated to the WM group were Community Schools (50% compared to 29% and 30% of those in the WM+ and Control group respectively). The WM+ and Control groups were more likely to be Academies (44% and 39% respectively, in comparison to 26% of those in WM). The proportion of rural schools in the control group was smaller than that in WM and WM+ (7% versus 17% in each of the treatment arms).

While some of these characteristics might have benefited the WM group, the proportion of pupils receiving FSM was also higher in this group, at 39% of pupils in comparison to 30% in WM+ and 34% in the control group.

Table 3: Ofsted ratings for schools

	Intervention groups				Control group	
	Working Memory		Working Memory Plus			
School-level	n/N	Percentage	n/N	Percentage	n/N	Percentage
Outstanding	8/42	19%	4/41	10%	4/44	9%
Good	29/42	69%	34/41	83%	29/44	66%
Requires improvement	3/42	7%	3/41	7%	4/44	9%
Inadequate	1/42	2%	0/41	0%	1/44	2%
Unknown	1/42	2%	0/41	0%	6/44	14%

Table 4: School Type

	Intervention groups				Control group	
	Working Memory		Working Memory Plus			
School-level	n/N	Percentage	n/N	Percentage	n/N	Percentage
Community School	21/42	50%	12/41	29%	13/44	30%
Academy	11/42	26%	18/41	44%	17/44	39%
Voluntary Controlled/aided school	8/42	19%	8/41	20%	11/44	25%
Other	2/42	5%	3/41	7%	2/44	5%
Unknown	0/42	0%	0/42	0%	1/44	2%

Table 5: Pupils eligible for FSM

	Intervention groups				Control group	
	Working Memory		Working Memory Plus			
School-level	n/N	Percentage	n/N	Percentage	n/N	Percentage
>50%	0/42	0%	0/41	0%	1/44	2%
25% - 50%	6/42	14%	8/41	20%	7/44	16%
<25%	34/42	81%	29/41	70%	29/44	66%
Unknown	2/42	5%	4/41	10%	7 /44	16%

Table 6: School Location

	Intervention groups				Control group	
	Working Memory		Working Memory Plus			
School-level	n/N	Percentage	n/N	Percentage	n/N	Percentage
Urban	35/42	83%	34/41	83%	40/44	91%
Rural	7/42	17%	7/41	17%	3/44	7%
Unknown	0/42	0%	0/41	0%	1/44	2%

Prior to randomisation, the project team identified two schools who shared a SENCO and two that were federated. The final number of units for randomisation thus decreased from 127 to 125.

Schools were asked to nominate children suitable for the trial. The criterion given to schools to nominate their children specified that children should be in the lower third or lower half of the year group, depending on the number of children in the cohort, at the end of KS1 in arithmetic. Schools had to have a minimum of 18 pupils in year 3 to participate. The guidance for the number of children taking part in the project is set out in table 7 below:

Table 7: Participation guidelines

No. of children in Year 3 (2016/17)	No. of children able to take part in the project
Less than 20	9
20 - 30	10
31 - 59	15
60 - 150	20

The Oxford University team sought opt-out consent for data sharing. Parents (or legal guardians) were sent an information sheet setting out details of the aims of the research. The form also offered parents the opportunity to opt out of the trial. The same form offered the opportunity for parents to separately withhold consent to accessing their child(ren)'s National Pupil Database (NPD) records. During the initial stage of data collection in which participating pupils' data were submitted, schools were advised to leave the Unique Pupil Number (UPN) field blank where pupils' parents had opted out of releasing this data. Of the 1475 randomised, 1246 pupils' records were submitted to the NPD, following the exclusion of 229 pupils, for whom the school had not provided their UPN.

Of the 1475 pupils finally randomised into the trial, 481 were assigned to the WM intervention, and 482 were assigned to the WM+ intervention. The remaining 512 pupils were allocated to the control arm. Of the total number of pupils randomised into the trial, there were a number for whom it was not possible to collect complete data. These are detailed in Table 8 below, with reported reasons:

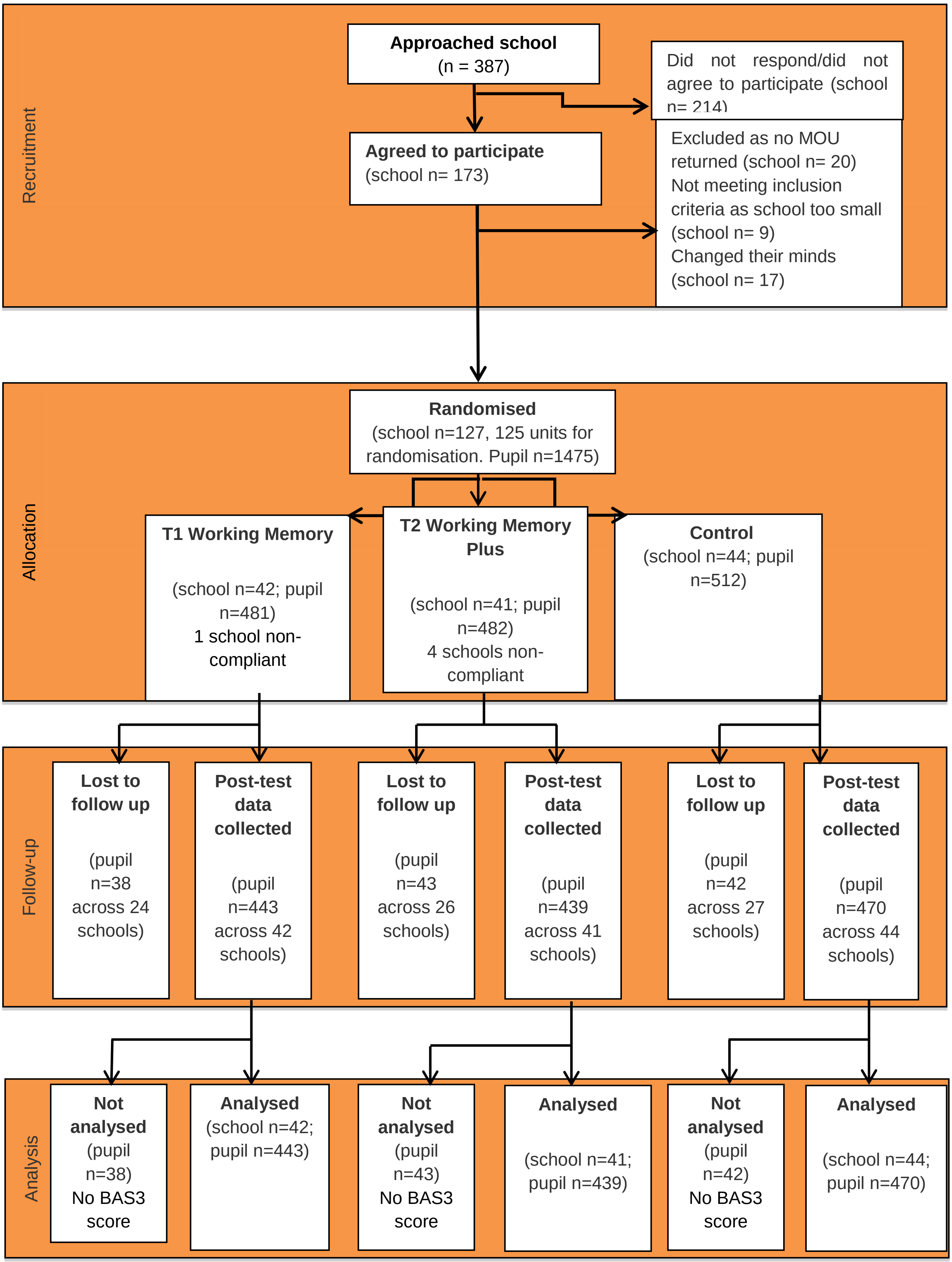
Table 8: Data Collection

Data	n/N (Proportion of sample missing)	Reason
KS1 Arithmetic Scores	77/1475 (5%)	Schools were not able to provide these data, and raw scores are no longer reported in the NPD
KS1 Reasoning Scores	87/1475 (6%)	Schools were not able to provide these data, and raw scores are no longer reported in the NPD
FSM, SEN, Absence	229/1475 (16%)	UPN consent was not provided to enable data to be pulled from the NPD
IDACI (Income Deprivation Affecting Children Index)	2/1475 (1%)	The NPD could not find a match for these UPNS
BAS3 Score	123/1475 (8%)	Pupils were absent on the day of testing, or had left the school
Attention and Behaviour Scores	78/1475 (5%)	Pupils were absent on the day of testing, or had left the school
Working Memory Scores	60/1475 (4%)	Pupils were absent on the day of testing, or had left the school

Figure 1 below provides details for the flow of participants through the trial. Following randomisation, one school withdrew, however still agreed to participate in testing, and was thus deemed as non-compliant.
None of the schools that were pre-tested were lost to follow up, so there was no school level attrition. There were, however, five schools deemed non-compliant, as due to other pressures they were not able to implement the programme. These schools all agreed to post-testing.

The pre-tests and post-tests were delivered during school hours, and thus some pupil level attrition occurred due to absence on the day of testing. Of the total sample, 6% of pupils were missing either both, or one of their baseline KS1 and KS2 scores. 8% of pupils were absent during data collection for the primary outcome measure, and as such these pupils were also excluded from analysis. Missing KS1 baseline scores were imputed for the subset of pupils for whom we had outcome data, and as such are not reflected in the diagram below. Pupils who were missing KS1 data due to their missing UPNs were still included in the primary analysis provided BAS3 outcome data had been collected for them.

Figure 1: Participant flow diagram



The difference between the assumed numbers of schools and children and the numbers recruited to the trial alters the power of the analysis. Table 9 presents this in the form of the minimum detectable effect size (MDES). The MDES is the smallest impact that the trial can reasonably be expected to be sensitive enough to register. It is measured in units of the standard deviation of the outcome. Following the convention of 80% power and 95% significance, the MDES reported in the protocol is 0.18. This is set out in table 9 below, along with the other assumptions used. As noted already, the target of 115 schools was exceeded and 125 schools were randomised (this includes the two school-pairs discussed earlier). The power gain achieved by this increase was offset by the average number of pupils per school being smaller than that assumed by the protocol, with the net effect of leaving the MDES unchanged. The bottom row of Table 9 presents the MDES for the analysis sample. In addition to reflecting the number of schools and pupils on which impacts are based, the observed ICC and correlation between regressors and the post-test can now be included. The ICC is higher than that assumed at the design stage (0.16 compared to 0.12) and the correlation is lower (0.50 compared to 0.75, partly due to the missing KS1 scores). Together, these have the effect of increasing the MDES to 0.26.

Table 9: Minimum detectable effect size at different stages

Stage	N schools (n=WM; n=WMp; =control)	Correlation between pre-test (+other covariates) & post-test	ICC	Blocking/ stratification or pair matching	Power	Alpha	Minimum detectable effect size (MDES)
Protocol	115 (38;38;37)	0.75	0.12	6 blocks defined on basis of school size and attainment	80%	0.05	0.18
Randomisation	125 (42; 41; 42)	0.75	0.12	6 blocks defined on basis of school size and attainment	80%	0.05	0.19
Analysis (i.e. available post-test)	125 (42; 41; 42)	0.50	0.16	6 blocks defined on basis of school size and attainment	80%	0.05	0.26

3.2 Pupil characteristics

Demographic data is presented in table 10 below, with all figures rounded to one decimal place. Whilst we expect no systematic bias to have arisen at the point of randomisation, attrition and missing data may create imbalance across trial arms.

Some differences were revealed during balance checks on KS1 baseline variables. KS1 arithmetic scores were higher on average for control arm pupils at 10.9, compared to a mean score of 10.6 in each of the treatment groups. KS1 reasoning scores were higher on average for WM pupils, who scored 13.5 average, compared to mean scores of 13.3 in each of the other groups. The highest degree of imbalance on this measure was 0.05 standard deviations. Histograms for these scores can be found in Appendix F. Similarly, there was some imbalance in scores for each of the secondary outcome pre-test measures, see Table 11.

In terms of pupils ever eligible for FSM, the WM group had a greater proportion in comparison to control, and control had a greater proportion in the WM+ group. The imbalance was 0.09 standard deviations. Analysis was thus repeated, controlling for FSM status. The results of this are discussed below.

Table 10: Balance checks on pupil characteristics

Variable	Intervention groups				Control group	
	Working Memory		Working Memory Plus			
Pupil-level (categorical)	n/N (missing)	Percentage	n/N (missing)	Percentage	n/N (missing)	Percentage
Gender: Female	481/1475 (0)	47.8%	482/1475 (0)	50.6%	512/1475 (0)	49.0%
FSM Eligible	420/1475 (61)	38.6%	380/1475 (102)	29.7%	444/1475 (68)	34.0%
Ethnicity: White (vs Non-white)	420/1475 (61)	82.4%	380/1475 (102)	76.1%	444/1475 (68)	77.7%
SEN (without Statement)	420/1475 (61)	27.6%	380/1475 (102)	29.2%	444/1475 (68)	31.3%
EAL	420/1475 (61)	14.8%	380/1475 (102)	17.9%	444/1475 (68)	17.3%
Pupil-level (continuous)	n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)
IDACI	420/1475 (61)	0.22 (0.14)	380/1475 (102)	0.19 (0.14)	442/1475 (70)	0.20 (0.14)
Absence (av. number of sessions missed during the year))	420/1475 (61)	11.5 (11.6)	380/1475 (102)	10.1 (10.2)	444/1475 (68)	10.8 (12.6)

Table 11. Balance checks on Pre-test Scores (control arm as reference group)

	Working Memory			Working Memory Plus			Control Group	
Pupil-level (continuous) Pre-tests	n/N (missing)	Mean (SD)	Effect Size (95% CI)	n/N (missing)	Mean (SD)	Effect Size (95% CI)	n/N (missing)	Mean (SD)
KS1 Pre-test score: Arithmetic	473/1475 (8)	10.6 (5.87)	0.04 (-0.95, 0.17)	435/1475 (47)	10.6 (6.13)	0.05 (-0.09, 0.19)	490/1475 (22)	10.9 (6.38)
KS1 Pre-test score: Reasoning	464/1475 (17)	13.5 (6.89)	-0.05 (-0.19, 0.09)	435/1475 (47)	13.3 (7.10)	0.01 (-0.16, 0.13)	489/1475 (23)	13.3 (7.66)
Listening Recall	476/1475 (5)	8.1 (3.06)	-0.09 (-0.21, 0.04)	479/1475 (3)	8.3 (2.94)	-0.15 (0.28, 0.03)	505/1475 (7)	7.8 (3.04)
Counting Recall	476/1475 (5)	13.7 (3.51)	-0.10 (-0.23, 0.02)	478/1475 (4)	13.6 (3.40)	-0.06 (-0.19, 0.06)	504/1475 (8)	13.3 (3.53)
Backward Digit Recall	476/1475 (5)	9.1 (2.85)	-0.06 (-0.19, 0.06)	478/1475 (4)	9.0 (2.63)	-0.02 (-0.14, 0.11)	505/1475 (7)	9.0 (2.63)
Attention and Behaviour	478/1475 (3)	32.0 (11.32)	-0.13 (-0.26, -0.00)	475/1475 (7)	32.3 (11.51)	-0.16 (-0.29, -0.03)	498/1475 (14)	30.5 (11.52)

3.3 Outcomes and analysis

Table 12 presents the results of the analysis. Effect sizes are shown for WM vs. control and WM+ vs. control. We do not estimate or report the effect size of WM+ vs. WM, because the trial was powered to detect expected effect sizes between WM/WM+ arms and control arms, while the expected effect size is much smaller when comparing the active trial arms and the analysis would be underpowered. Effect sizes were calculated using Hedges-g.

For the primary outcome, number skills as measured by BAS3, WM shows a positive effect of 0.19 standard deviations (-0.10, 0.47) and WM+ shows a positive effect of 0.24 standard deviations (-0.05, 0.52). In both cases, the confidence intervals span zero, suggesting that we cannot discount the possibility that these differences merely reflect chance variation. On this basis, we conclude that there was no statistically significant effect on BAS3 number skills. Histograms for each group are provided in Appendix F. We note that the BAS3 scores for each group are skewed, with a significant proportion of pupils scoring at or near the limit for this test. While a ceiling effect is present, the movement of scores around a density of .3 across trial arms indicate that there is still sensitivity in this measure for the top end of the distribution. Given the intervention is designed to target those pupils performing below the class average in numeracy, the instrument is sufficiently sensitive to capture changes for the target group.

As a robustness check, we also conducted the analysis using both null imputation and the alternative, with no imputation of the baseline covariates. Under null imputation, missing values are set to zero and additional variables are included identifying pupils for whom this change has been made. The resulting effect sizes are comparable to those already presented (based on multiple imputation). WM is estimated to increase number skills by 0.16 standard deviations and WM+ is estimated to increase number skills by 0.26 standard deviations. Relative to the multiple imputation results, the estimates appear marginally more precise. However, we view the multiple imputation results as better capturing the uncertainty around missing values and consequently regard those as the preferred results.

Using no imputation, estimates are based on all pupils for whom baseline data and BAS3 endline data are not missing. While the effect sizes for WM and WM+ (0.18 and 0.27 respectively) remain similar to those observed above, the level of statistical significance for the WM+ arm increases marginally to reach the 95% level.

The results of supplementary analysis in which WM and WM+ are pooled reflect the findings of the original analysis, in which the effect for both arms are not significant. Results for these robustness checks are available in Appendix E. This was not set out in the SAP, but was undertaken following suggestions from an independent peer reviewer.

Table 12 also presents estimates of the impact on secondary outcomes. These are more closely aligned with the content and aims of the intervention and so one might expect them to be more likely to register a significant effect. This is indeed the case.

For listening recall, counting recall and backward digit recall, WM had statistically significant effect sizes of 0.43, 0.76 and 0.56 respectively. Interestingly, the corresponding WM+ effect sizes were slightly smaller (0.18, 0.61 and 0.27 respectively) and, in the case of listening recall and backward digit recall, fell short of statistical significance at the conventional level. The recall factor score provides an estimate of the impact across all three recall outcomes. This confirms the positive effect of both WM and WM+ and suggests the stronger effect of the former. On this measure, WM increased recall by 0.37 standard deviations and WM+ increased recall by 0.22 standard deviations.

Table 12 also presents the estimated impact on attention and behaviour in class. This is coded such that a negative effect represents an improvement. The results indicate the programme was associated with better attention and behaviour in class. This was particularly the case with WM+ which shows a significant effect size of 0.69. For WM, the effect size is slightly smaller (0.44), so suggestive of an improvement albeit this falls short of the conventional threshold for statistical significance.

Lastly, we examine the estimated impact for pupils ever eligible for FSM. The estimates appear smaller than those for the full sample, but are similar in not being significantly different from zero.

Supplementary analysis to control for FSM in the model show that the effects for both arms remain non-significant. These results are presented in Appendix E. This was not set out in the SAP, but was included following recommendations from an independent peer reviewer.

Table 12: Impact estimates

	Raw means						Effect size		
	WM group		WM+ group		Control group			WM group	WM+ group
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (WM; WM+; control)	Hedges g (95% CI) p-value	Hedges g (95% CI) p-value
Primary outcome									
BAS3 score with multiple imputation	481 (38)	8.8 (3.3, 14.3)	482 (43)	8.8 (3.3, 14.3)	512 (42)	8.5 (2.6, 14.4)	1352 (443; 439; 470)	0.19 (-0.10, 0.47) 0.20	0.24 (-0.05, 0.52) 0.11
BAS3 Score (FSM subgroup)	162 (16)	8.3 (2.7, 13.9)	113 (9)	8.1 (2.0, 14.1)	151 (16)	8.0 (2.1, 13.9)	385 (146; 104; 135)	0.09 (-0.14, 0.32) 0.49	-0.02 (-0.28, 0.23) 0.87
Secondary outcomes									
Listening recall	481 (21)	10.8 (4.4, 17.2)	482 (22)	10.4 (4.3, 16.6)	512 (17)	9.9 (4.3, 15.5)	1415 (460; 460; 495)	0.43 (0.21, 0.66) 0.00	0.18 (-0.09, 0.44) 0.19
Counting recall	481 (21)	17.3 (8.2, 26.4)	482 (22)	16.8 (9.0, 24.7)	512 (17)	15.6 (7.7, 23.4)	1415 (460; 460; 495)	0.76 (0.41, 1.11) 0.00	0.61 (0.28, 0.93) 0.00
Backward digit recall	481 (21)	11.3 (3.3, 19.3)	482 (22)	10.7 (4.3, 17.0)	512 (17)	10.2 (4.0, 16.4)	1415 (460; 460; 495)	0.56 (0.28, 0.84) 0.00	0.27 (-0.02, 0.55) 0.07
Recall factor score	481 (21)	0.2 (-2.0, 2.4)	482 (22)	0.0 (-1.8, 1.8)	512 (17)	-0.2 (-2.0, 1.5)	1415 (460; 460; 495)	0.37 (0.22, 0.52) 0.00	0.22 (0.06, 0.37) 0.01
Attention and Behaviour	481 (26)	28.8 (9.2, 48.3)	482 (23)	28.0 (7.2, 48.9)	512 (29)	29.3 (8.8, 49.9)	1397 (455; 459; 483)	-0.44 (-0.92, 0.05) 0.08	-0.69 (-1.25, -0.14) 0.01

Game Data Analysis

As noted above, the intervention involved significant playing of games as part of the independent technique practice. 30 minutes of time for playing games were included in the sessions. As such, variation in game-playing may provide us with some measure of intervention dosage within the treatment groups.

There was significant variation in the number of games played both between individuals and between schools taking part in the project. The mean number of games played by an individual was 37, with the median also being 37. Some children played only a single game, while others played more than eighty.

Figure 2: Histogram of number of games played per pupil

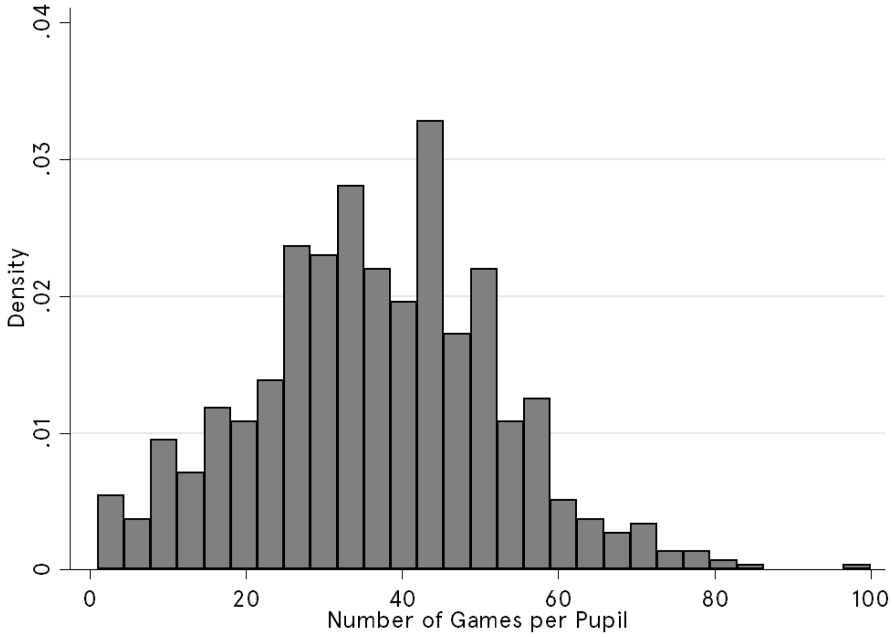
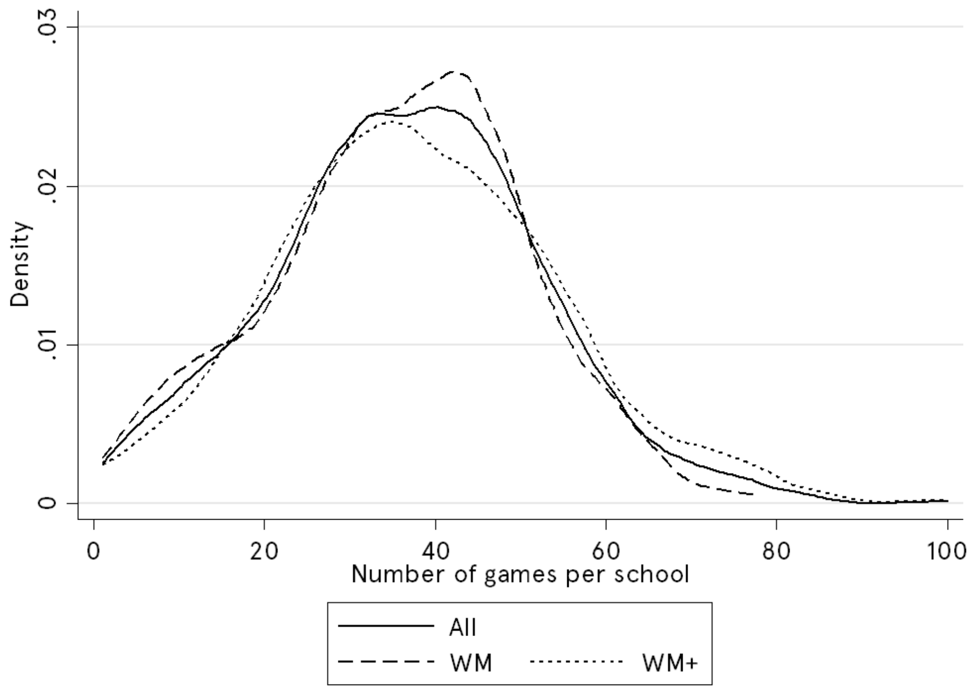


Figure 3: Kernel density plot of number of games played per pupil by treatment arm



Between schools, the mean number of games played in a school was 407, while the median is 402 (reflecting an upper tail of a small number of schools in which a particularly high number of games was played).

Figure 4: Histogram of number of games played per school

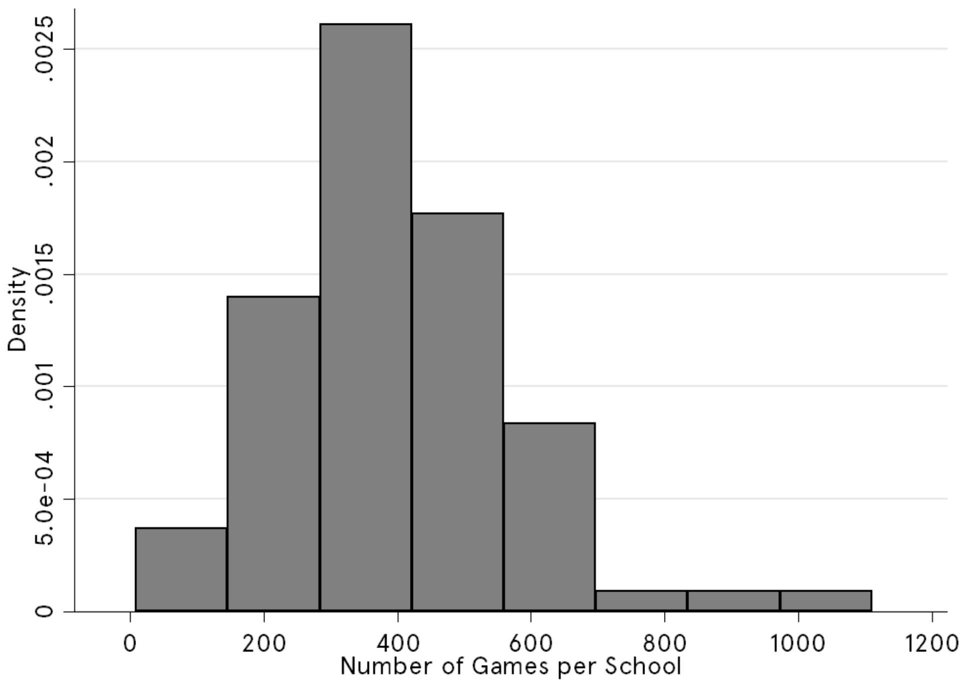
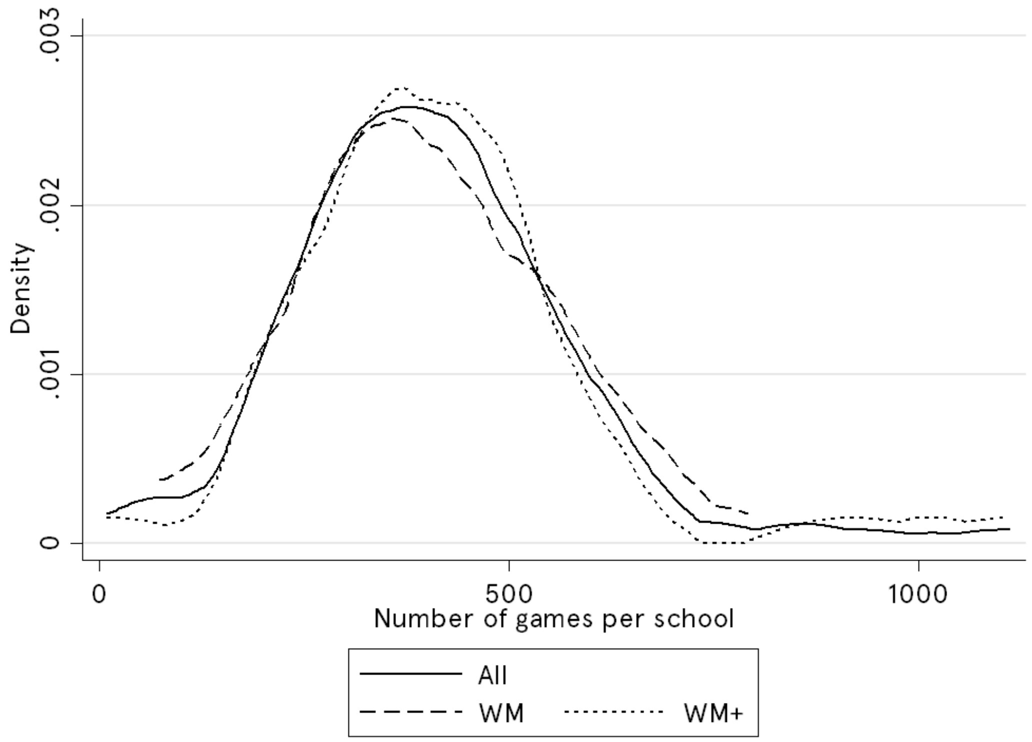


Figure 5: Kernel density plot of number of games played per school, by treatment arm



In order to understand whether the differences in numbers of games played by treatment arm is statistically significant, we use a linear regression model to test difference in means. The first column Table 13 demonstrates that there is slight evidence of a higher number of games being played by pupils in schools allocated to the WM+ arm (36 games per pupil, compared with 34 games per pupil). However, this difference is not statistically significant once we take into account the school-level clustering. Similarly, the second column of Table 13 demonstrates that there is slight evidence of a higher number of games being played in a school allocated to the WM+ arm (397 games per school, compared with 375 games per school). Again, this difference is not statistically significant.

Table 13. Linear regression of number of games played by treatment arm

	Games Played	Games Played
	Individual	School
Working Memory Plus Arm	1.78	21.64
	(0.67)	(0.55)
Constant	34.43***	375.24***
	(8.52)	(6.12)
N	866	78

Notes: t statistics based on school-level clustered standard errors reported in parentheses. Stars indicate statistical significance as follows: * p<0.05; ** p<0.01; *** p<0.001.

Table 14 shows that pupils' scores tended to decline as they did more games (column 1), however this seems to reflect the fact that they were moving onto harder games. After including a game-type fixed effect to control for this (column 2), this reverses and we see that pupils' performance in a given type of game improves as they do more of them. The increase is even stronger once we include a game-specific fixed effect (column 3), effectively comparing what happens when pupils do the same game more than once.

Table 14: Linear regression of game score by number of games played

	Game Score	Game Score	Game Score
	(1)	(2)	(3)
Game Number	-0.358***	0.115**	0.297***
	(-6.99)	(3.10)	(4.96)
Constant	95.44***	82.34***	71.97***
	(113.79)	(50.91)	(25.20)
Game type indicator		√	
Game-specific indicator			√
N	31,757	31,757	31,757

Notes: t statistics based on school-level clustered standard errors reported in parentheses. Sample restricted to individuals with reported outcomes data. Stars indicate statistical significance as follows: * p<0.05; ** p<0.01; *** p<0.001.

Finally, we explore whether there is evidence of differences in the number of TA sessions or Computer Games sessions it is reported that pupils took part in. Pupils should have participated in 10 TA sessions and 10 computer games sessions.

Table 15. Distribution of TA Sessions in which pupils participated by treatment arm.

Number of sessions	WM (%)	WM+ (%)	Total (%)
9 or fewer	46	17	33
10 or more	54	83	67
Total	100	100	100

Notes: Reporting column percentages. Categories have been collapsed due to small cell sizes. N=838.

Table 16. Distribution of Computer Games sessions in which pupils participated by treatment arm.

Number of sessions	WM (%)	WM+ (%)	Total (%)
9 or fewer	33	20	27
10 or more	67	80	73
Total	100	100	100

Notes: Reporting column percentages. Categories have been collapsed due to small cell sizes. N=851

The distribution of number of TA sessions in which they have participated is reported in Table 15, while the distribution of number of Computer Games sessions in which they have participated is reported in Table 16. The median number of sessions of both types in which pupils participate is 10, while the mean in both cases is closer to 9, due to the pupils who received fewer than their full allocation. It is easier to interpret this in terms of the distribution taking different numbers.

Approximately two-thirds of pupils received the full 10 TA sessions (in fact, in the WM+ arm, it appears a small number of pupils received up to 15 TA sessions; precise details not shown due to small cell sizes). It is a similar picture for the Computer Games sessions with 73% of pupils received the full 10 planned.

Only just over half of pupils in the WM arm received the full 10 TA sessions, while 80% of those in the WM+ arm did so. Likewise, two-thirds of pupils in the WM arm received the full 10 Computer Games sessions, while almost 80% of those in the WM+ arm did so.

3.4 Cost

The cost per pupil per year over 3 years is estimated to be £22.59 per pupil in the WM arm, and £23.72 in the WM+ arm. In each case, start-up costs per school, running costs and running costs per pupil were specified and included in estimates. These are summarised in table 17 below.

Table 17. Cost calculations

	Running cost per pupil	Start-up cost per school	Total cost over 3 years	Cost per year over 3 years	Cost per pupil per year
Working Memory	£0.53	£794.33	£813.41	£271.14	£22.59
Working Memory Plus	£1.37	£804.63	£853.95	£284.65	£23.72

The cost of two laptops at £360 each was the only recorded start-up cost per school, as schools were advised that in order to deliver the intervention, staff would need access to two computers or laptops. The majority of schools purchased this equipment specifically for the purpose of delivering the intervention.

Subsistence costs were included, at between £14 - £25 per school. These consisted of all costs associated with travel to and from the training day for the two members of school staff in attendance. As staff training only occurs once, this is considered a start-up cost.

Resource costs per pupil over three years were included to arrive at the final total per pupil estimate for each arm of the trial over 3 years. Costs for resources (printing and stationery) are estimated at around £6 - £17 for a year per school. Teachers and TAs were advised to print record sheets for each pupil, which formed the bulk of the photocopying and printing costs for the intervention.

Training and teacher cover

The schools in each trial had either TAs or teachers allocated to support the delivery of the intervention. Each was required to attend one training day outside of the school. In many cases, where teachers were sent to attend the training day, schools employed TAs at additional cost to cover the days these teachers were absent. In terms of staff time, training is estimated at 8 hours per teacher/TA, and a total of 16 hours per school, given two members of staff were usually sent to attend training. Schools who were allocated to receive the WM+ intervention reported that they were required to spend approximately half a day extra in training, to be instructed in the delivery of the extra games required. This increases training time for two members of staff to approximately 24 hours. Where schools were required to arrange cover for absent teachers, their time was estimated at the same number of hours required for the training. Maximum staff time is therefore estimated at 32 hrs for the WM arm, and 48 hours for the WM+ arm, including teacher cover.

Preparation

There were some additional staffing costs associated with preparation for the administration of the intervention, which varied across schools. This amounted to approximately four hours of preparation before delivery began (to allow for meetings), and five hours over the course of the intervention (for preparation prior to each session). Assuming some administration is required (irrespective of whether the project is a trial), staff time for preparation over three years would amount to approximately 27 hours.

Delivery

The estimated minimum number of hours of staff time required for the delivery of the intervention over the course of the year is one hour per pupil. Given the average size of the group treated in this trial, for a class of 12 pupils, we estimate 120 hours of staff time for one year, and 360 hours for three years.

Run as a trial, there were also costs associated with the time taken for a member of staff to oversee the intervention and for senior leadership time. This ranged from up to 30 mins per week, to a total of 7 hours for time spent speaking to the evaluation team and project team. These costs are not included in the intervention costs per school and per pupil. Excluding this time, for a school administering the intervention to 12 pupils over three years, total staff time is estimated at 419 hours, or approximately 52 days. Given the extra training required for teachers and TAs in schools allocated to the WM+ arm, and we estimate a maximum of 54 days total staff time for these schools.

Process evaluation

The findings presented in this section are based on the qualitative data collected via visits to eight treatment schools and by surveying treatment and control schools. It should be noted that the views and experiences presented here are not necessarily representative of all treatment and control schools. The end of project survey was completed by 74 per cent of all participating schools (total N=127). Further, while steps were taken to ensure that the treatment schools visited as part of fieldwork included a variety of delivery contexts, the number that participated in this part of the research was very small compared to the larger treatment population. As such, some perspectives on the delivery of the improving working memory project may be missing from the final analysis.

In the discussion that follows, it should be noted that survey responses are only presented in percentages when the total number of respondents/responses is equal to or greater than 100.

4.1 Perceived need for intervention

Participants' experiences of the early stages of implementation were explored through the interviews and online survey of link teachers and TAs involved in delivery within treatment schools. Many interviewees commented that they saw a clear need for the intervention within their school. Specifically, several link teachers perceived that the current cohort of year 3 pupils struggled to retain information and recall what they had covered in previous lessons, particularly in maths. The focus of the project on improving pupils' working memory skills was therefore seen to be potentially helpful in addressing this issue:

'In the class as a whole, things like mathematical skills, you'll come back to them and the children do have very low retention [..,] So that was where we hoped that perhaps we'd be able to see a difference made'.

Link Teacher, School 8, WM group

'With the maths, in particular, we're really concerned about how much they're retaining information – with times tables, number bonds and things like that, it just literally goes. Some of mine in my class, I'm asking them what numbers make 10 – and I know they've done it in year 1 and year 2, but they just can't seem to retain it. So, I think we just wanted to help with the whole memory thing, and taking in more information'

Link Teacher, School 1, WM+ group

Other link teachers related the school's decision to become involved in the project to its poor performance in maths across year groups and the deficiencies in pupils' ability to recall information that this exposed. In these cases, the schools concerned wanted to trial this approach, see whether it was effective and potentially cascade it to pupils in other year groups within the school.

In spite of this widespread acknowledgement among interviewees of the potential benefits of the intervention, few had ever heard of the concept of working memory before. Where they had, interviewees spoke of having obtained a basic understanding from their education while one link teacher recalled having delivered some working memory games in the past for a pupil in their class who had severe dyslexia.

4.2 Experience of training and preparedness for delivering the intervention

All treatment schools across both treatment arms received 1 day of training prior to implementing the intervention. The link teachers as well as the TAs who would be delivering the sessions were asked to attend. Where TA's were not able to attend, training was provided for them at their school. During the training, the delivery team briefed attendees on the background to the project, introduced the concept of working memory and presented the research evidence from previous research on the effectiveness of this approach. A large part of the day, however, was spent introducing the computer-based activities and games that pupils would be completing with the TA and independently. Attendees were asked to bring a laptop to the training session, so they could access the online resources developed for this project and spend time practicing the delivery of the computer-based activities with support and guidance from the delivery team.

Those schools in the WM+ treatment arm of the trial received an additional half-day training session a few weeks after beginning delivery. This session introduced attendees to the two components of arithmetic knowledge that this part of the intervention is intended to promote: additive composition and the inverse relation between addition and subtraction. Attendees were again made familiar with the computer-based activities and games that they would be delivering for the final five weeks of implementation and had the opportunity to practice these within a supported environment.

In both training sessions, attendees received a pack of materials that included a research briefing on the core concepts covered by the intervention as well as a handbook that instructed TAs on how to complete the computer-based activities with pupils.

Overall, interviewees from treatment schools were very positive about their experience of attending the training sessions. In particular, both link teachers and TAs liked the practical aspect of the training. They commented that it was helpful to have the opportunity to rehearse delivering the computer-based activities and see how the independent games and platform they were based on worked. Several TAs mentioned that this made clear the practicalities of their role in the project, and allowed them to identify and resolve any uncertainties they had about how the activities and games were meant to be implemented:

‘We were all involved in the training session, we were able to actually work with another person and actually do that activity which was really helpful, because you could see in practice and see if anything arose from the session or the activity. And we were able to identify the problems and ask questions’

Teaching Assistant, School 4, WM+ group

Where some interviewees did not perceive the training session(s) to have been beneficial, this was due to difficulties in accessing the computer-based activities due to technological problems and not having brought a laptop to the training session.

With regards to the handbook that TAs received in preparation for implementing the intervention, interviewees were similarly positive commenting that it was an accessible and comprehensive guide for delivery.

Treatment schools were asked how prepared they felt for delivery following their attendance at the training day(s) and after reading through the accompanying handbook. The results of the online survey show that most of the TAs that responded to this question (33 out of 45) felt either well prepared or very well prepared for delivering the intervention, while 12 felt somewhat prepared. The interviews with TAs supported these findings and indicated that, where they did not experience any issues in accessing the computer-based activities during the training session(s), interviewees felt ready to deliver the intervention and needed limited amount of time afterwards to prepare. This involved putting together a folder of all the resources they would need, such as attendance sheets and progress charts.

In contrast, TAs who were unable to practice the computer-based activities at the training spoke of how they did not have clear idea, at this point, of how they were meant to implement the intervention and had to subsequently find time in their own schedule to rehearse the activities. For one treatment school, this delayed the start of delivery.

4.3 Number and type of pupils selected to participate

Prior to randomisation, schools were required to identify eligible pupils from the year 3 cohort for 2016/17 to participate in the trial. Specifically, schools were asked to select those pupils that had the lowest numeracy skills in the year group. They were advised to select a minimum of 10 pupils to participate, and larger schools with more than one year 3 form group were encouraged to select more (up to a maximum of 20 pupils). Schools were asked not to include pupils who had a SEN statement.

As shown in Table 18.1, the findings from the end of project survey, which are based on recall and perceptions of link teachers, indicate that most treatment schools that responded (54 out of 68) followed the instructions from the delivery team with regards to selecting pupils from the lower end of the class in terms of their ability in maths and other subjects. The fieldwork visits highlighted that most schools found it straightforward to identify pupils with the lowest number skills to participate in the intervention. However, in the case of one participating school that had small class sizes, the link teacher commented that it had been difficult to identify a minimum of 10 pupils who were working below the class average in maths. They had therefore chosen three more pupils to participate in the intervention who were working at the class average in order to meet this requirement.

Table 18.1 also shows that four respondents recalled selecting pupils at random to participate, contrary to the guidance from the delivery team. This included one school that the evaluation team visited as part of fieldwork for the process evaluation. The staff involved in the project explained that they selected every fifth child from the class register across three year 3 form groups to take part. They had felt that it would be interesting to see who the intervention worked best for across a range of ability levels and were not seemingly aware of the request from the delivery team to deliver the programme to pupils with the lowest number skills.

Table 18.1 – How pupils were selected to participate in the IWM programme

	Overall	WM	WM+
Pupils were selected from the lower end of the class	54	31	23
Pupils were selected at random	4	1	3
Don't know/Can't remember	10	2	8
N	68	34	34

The findings from the end of project survey indicate that the majority of treatment schools that responded (63 out of 69) recalled between 10 and 20 pupils starting the programme in their school, which is within the appropriate range specified by the delivery team. Forty-four respondents recalled the minimum requirement of 10 pupils starting the programme in their school. These findings differed little between treatment allocations.

During the fieldwork visits to treatment schools, it was common for staff to base their decision on how many pupils to include in the intervention on the anticipated impact it would have on internal resources. For instance, among schools that selected 10 pupils to participate, staff stated that they felt this was the maximum number they could deliver to given the time commitment it would require from TAs each week (i.e. five hours) in addition to the practicalities of finding a suitable quiet space and securing access to two computers for each session.

‘It is a big commitment, resource wise, and that’s why we decided to go with only ten children’.

Link Teacher, School 4, WM+ group

'It's time factors always in schools, and plus access to laptops and things because we haven't got that many and they are always in demand. So you know just practicalities of room space that was quiet enough for the children to work individually, and to have access to a laptop. We really couldn't have managed with any more children just from that point of view'.

Link Teacher, School 2, WM group

In contrast, staff from one school that had put forward 20 pupils to participate stated that this decision had been taken by a member of the school's senior leadership who was not attentive to what the project would require in terms of resource. The link teacher in this school reflected that they would have chosen to include fewer pupils in the programme had they had more knowledge of the nature of the intervention and what was needed from the school at the outset.

While most schools selected an adequate number of pupils to start the IWM programme, there was some attrition over the course of delivery. The end of project survey showed that fewer than half of schools that responded (29 out of 69) had 10 pupils complete the programme. Among those schools visited as part of fieldwork, dropout was common with most reporting that at least 1-2 pupils had withdrawn from the programme since it started. They indicated that this was most often due to pupils having left the school, and this was also reflected in feedback from respondents who completed the end of project survey.

As noted, the delivery team stipulated that pupils should not be included in the programme if they had a SEN statement. However, this did not preclude SEN pupils from participating who had not been statemented. The survey results show that a large number of treatment schools (43 out of 65) recalled that they had at least one SEN pupil receiving the intervention.

4.4 Length and number of sessions per week (dosage)

The intended dosage for the trial was for each child to participate in ten one-hour sessions; pupils would spend 30 minutes completing TA-led activities on the computer, and 30 minutes working independently playing a series of related online games. For each session, it was envisaged that TAs would bring two pupils out of class: one would complete the TA-led activities while the other would play the online games over the course of the hour, switching half-way through. For those in the WM group, all ten sessions focused on games that were intended to improve pupil's working memory. For the WM+ group, the first five sessions would focus on these games. However, after these were completed, pupils would switch to number-based games for sessions six to ten, which aimed to improve pupil's understanding of the arithmetic concepts of additive composition and the inverse relation between addition and subtraction. The findings presented in this section indicate that this intended dosage for the trial was largely met, at least among the treatment schools that responded to the survey.

As previously indicated, most treatment schools (44 out of 69 respondents) that participated in the end of project survey reported that 10 pupils started the programme in their school. In order to achieve the required dosage for this trial, therefore, the majority of schools would need to deliver five one-hour sessions each week with two pupils at a time. The survey findings indicate that this was partially achieved: 26 out of 46 TAs stated that they were able to deliver 5 sessions in a typical week over the course of the programme (although attrition will have affected the number of sessions TAs were typically able to deliver). They also tended to be the appropriate length, with 31 out of 47 respondents recalling that sessions typically lasted for 60 minutes.

The survey also asked TAs how often they were able to maintain the 50/50 spilt between the time pupils spent on the TA-led games compared to the online games over the course of a session. As Table 18.2 shows, 26 out of 47 TAs that responded to the survey felt that they were able to maintain this 50/50 split most of the time, while 11 stated that they were able to maintain the split all of the time.

It should be noted that slightly fewer respondents in the WM+ group stated that they typically felt able to maintain the 50/50 split between TA-led and online games over the course of a session, compared to those allocated to the WM group. During the fieldwork visits, some TAs in the WM+ group stated that the number-based games pupils completed during the latter half of the intervention were more challenging. As such, they found it more difficult to keep to time as pupils required greater support in order to make progress. While this experience did not appear to be widespread, it may partly explain the small differences in responses between treatment groups here.

Table 18.2 – How often TAs were typically able to maintain the 50/50 split between the time pupils spent on the TA-led games vs. the online games

	Overall	WM	WM+
All of the time	11	8	3
Most of the time	26	16	10
Some of the time	10	4	6
Hardly ever	0	0	0
Not at all	0	0	0
N	47	28	19

As well as the overall length and structure of the sessions, another important factor in considering whether the WM programme was implemented as planned is how easy or difficult it was for pupils to work independently through the online games for 30 minutes. The findings from the end of project survey indicate that TAs had varying experiences. Across treatment groups, 25 out of 47 TAs perceived that pupils had generally found it difficult to work independently. Eighteen respondents offered a neutral response stating that it was neither particularly easy nor difficult for them, while just four respondents stated that pupils had found it easy to work independently. TAs were subsequently asked to explain their responses to this question.

Pupils' difficulties in working independently were attributed to behavioural issues as well as a lack of IT skills. For those TAs who were neutral about how easy or difficult pupils found the independent games, it appeared that only a few pupils receiving the intervention had behavioural issues that caused them to lose concentration and become distracted over the course of 30 minutes. Further, no respondents in this group spoke of pupils not having the appropriate IT skills to access and navigate the online games.

Further detail on these issues came from the fieldwork visits. With regards to the behavioural issues some pupils exhibited, several TAs involved in implementation explained that these individuals required a lot of support and lacked confidence in their own abilities. As such, while they worked well on a one-to-one basis during the TA-led activities, they struggled to independently apply the strategies from these exercises to the online games and could become frustrated and disengage altogether when they were unable to obtain the correct answer. This behaviour was also noted in a few of the session observations. Some staff therefore reflected that 30 minutes was too long for these pupils to work on their own with little guidance or instruction, and that it was common for them to lose focus over a short period.

'It depended which children you had because we had a few with behavioural issues which was really difficult because they were fine when they're sat with you on that one-to-one basis, but then when you put them onto the laptop on their own that's when they couldn't focus and every two minutes it was, 'can you help me with this?' After a couple of sessions, I told them they must not disturb me, yeah unless it was absolutely urgent, but those children with the behaviour problems as well they found that so difficult to actually sit there for that length of time as well'.

Teaching Assistant, School 2, WM group

In several cases, pupils receiving the intervention had special educational needs such as dyslexia or attention deficit hyperactivity disorder (ADHD), which a few TAs noted could present an additional barrier to pupils concentrating independently for 30 minutes and not requiring support.

Some TAs mentioned that these issues were mitigated, to some extent, if pupils completed the TA-led activities at the beginning of a session and then switched to the online games. They mentioned that this helped pupils to recall how they were meant to complete each game and the appropriate rehearsal strategies and techniques they should be using. If they made progress in the TA-led activities, this was also seen to boost their confidence in their ability to complete the games: an attitude that they then transferred to independent working. However, one TA commented that they did not feel it was always possible to work with pupils who found the independent games difficult at the beginning of each session as this was unfair on the pupils who they were paired with who wanted some variation.

In terms of IT issues, a number of TAs commented as part of the survey and fieldwork visits that a lot of time was initially needed to explain and remind pupils of how to login to and play the online games. However, several noted that this did improve over time as pupils gradually learnt how to use the online resource. Few schools visited as part of fieldwork had pupils taking part that did not have the necessary IT skills to use a computer and access the online games. Where this did occur, the pupils concerned had additional learning requirements that presented them with difficulties when using a keyboard, for instance.

Another common IT issue that pupils encountered was that after completing a particular set of online games and answering most of the questions correctly they would not progress to either the reward games or navigation screen as expected. This problem occurred during a few of the session observations and was noted by several survey respondents. As TAs were not closely monitoring pupils' progress through the online games, they remarked that it was difficult for them to identify the root cause of the issue and prevent it happening in future. Where this occurred, pupils could become demotivated as their progress was not formally recognised on the system, and in a few cases they had to repeat the same exercises again. The delivery team and some TAs later determined that, in these instances, pupils may have not clicked on the correct links when progressing through the questions for a certain set of games.

'We had the odd occasion whereby children successfully completed all questions in a level but weren't automatically moved up a level. We believe this was because they may have skipped clicking on the "OK" or "next question" links at some point'.

Survey respondent, WM group

'Sometimes the children would get to the end of a game independently and it doesn't take them back to, for example, the first page, and then I couldn't identify, even though they told us, it could be this or that, but I couldn't identify the root cause of that issue; that happened a few times. But because we're not watching them continually, we can't determine the root cause'

Teaching Assistant, School 4, WM+ group

Other issues noted included pupils progressing and encountering new games while working independently, which they had not completed with the TA before. A few TAs noted that pupils found the games difficult to complete without any instruction about how the game worked and the appropriate strategies/techniques they should be using, as well as the chance to work through some examples with the TA first (i.e. guided practice). This occurred in several of the session observations. In these instances, the TA in question had to decide whether to interrupt their work with the other child to explain the rules of the new game or continue with the TA-led activities so as to not disrupt the session. Pupils completing the online games generally did not progress well where this happened, and even where TAs chose to take time to explain the rules of the new game to the child, they tended to answer most of the questions incorrectly without the opportunity to go through a few of the questions with the TA first.

Pupils' ability to work independently and concentrate for 30 minutes could also be affected by the physical environment they were in; if schools were unable to secure a quiet space to deliver the sessions then pupils could become easily distracted and make little progress through the online games.

4.5 Pupil responsiveness

The process evaluation also explored the extent to which pupils were engaged with and enjoyed the intervention, which helps to provide an indication of how responsive they were to its content. For instance, TAs were asked in the end of the project survey to rate pupils' level of overall enjoyment of the TA-led activities and the independent online games that aimed to improve pupils' working memory. TAs within the WM+ group were also asked to rate pupils' level of enjoyment of the arithmetic games in this arm of the trial.

For the working memory games, most respondents to this question (28 out of 46) perceived pupils' level of enjoyment of them as high or very high, while 17 stated that it was neutral. The interviews completed with staff in treatment schools supported these findings. Several interviewees observed that it had not been difficult to motivate pupils to engage with the programme, and they generally wanted to come out of class and enjoyed completing the working memory games. Pupils were enthusiastic that they had been selected to come out of lessons and receive this individual support; it made them feel *'special'* and caused several to *'raise their game'* when it came to completing the activities. In terms of the TA-led activities, TAs recalled that pupils generally liked the 'Colours' and 'Digits' games best as they found these exercises the easiest and made good progress through the various levels. 'Words' was seen to be more challenging, however, with some pupils progressing far slower and getting stuck on certain levels. This was seen to affect their engagement, as pupils became bored from frequently having to repeat the same games (and alternative games) and were demotivated by their lack of progress.

For the arithmetic games, the results were similar, though the total number of responses was very low (N=16). Ten out of 16 respondents perceived pupils' enjoyment of the games as being high or very high. During the fieldwork visits, some interviewees commented that pupils seemed to prefer the arithmetic games, as they were less repetitive and offered greater variety than the working memory games. Others stated that pupils were ready for a change in approach after five weeks and the transition to the arithmetic games renewed their interest in the programme. This challenged the preconceptions of some TAs who had believed that pupils would not be engaged in this part of the trial due to their dislike of maths as a subject. They commented that the fact that pupils did not lose interest was indicative of the quality of the programme and of the efforts made by the delivery team to make the games enjoyable.

However, as noted previously, some TAs did perceive the arithmetic games to be more challenging for pupils, with a few stating that some of the activities were at too high a level for pupils to complete. In particular, pupils were seen to struggle with some of the inversion games and a few took a while to make progress. Interviewees recalled how pupils struggled to grasp the concept of focusing on the relation between the number being added and the number being subtracted to arrive at an answer, as opposed to trying to complete the entire calculation in sequence, which some pupils persisted in doing.

The ‘Just numbers’ exercise was seen to be particularly difficult for pupils given the absence of any visual cues on the amounts being added and subtracted to assist them and also because it was a timed exercise.

A few schools, however, attributed these issues to their choice of pupils to take part in the project, as opposed to it being a fault with the programme. Staff from these settings stated that they felt that they had chosen pupils with numbers skills that were too low for this project. Some had individuals with special educational needs, such as dyslexia, in their group who struggled with letter and number recognition, and upon reflection they did not feel that the games were appropriate for pupils that had these additional requirements.

4.6 Adaptations to the intervention

The interviews with staff and observations of sessions indicate that treatment schools did not make any significant adaptations to the intervention. TAs spoke of how they strictly adhered to the handbook instructions provided by the delivery team, which specified how to complete the TA-led activities with pupils. In the observations of the working memory games, as per the handbook, all TAs were teaching and prompting pupils to use the appropriate oral rehearsal strategies and spatial cues for each game to support recall. This was done every time pupils started a new set of games, or when they began to struggle with their recall/provided an incorrect answer; TAs would practice the appropriate strategy with pupils until they were adopting it themselves. Similarly, as per the handbook instructions, when pupils provided the correct response to a particular game (in the case of both working memory and arithmetic), TAs would always ask pupils to explain how they had arrived at that answer at least for the first few exercises until they were happy with the pupil’s comprehension.

A few TAs mentioned that they occasionally altered certain words in the script if they seemed to be confusing pupils, while some TAs taught pupils using whatever phrasing felt most natural to them once they had a developed a good understanding of how the TA-led activities and online games worked.

4.7 Implementation environment

A few factors in the school environment were seen to affect how the intervention was delivered by staff. This included when the sessions were timetabled to take place, and whether TAs were able to secure the necessary resources needed to deliver the intervention: namely, two working computers with internet access and a quiet teaching area.

With regards to when the sessions took place, 34 out of 64 treatment schools that responded to the end of project survey reported that they were timetabled either exclusively or mostly in the afternoon. Staff explained during the fieldwork visits that this was to ensure that pupils did not miss core subjects (i.e. English and maths), which were taught in the mornings. As a result, the survey results show that pupils were most often taken out of PE, history, science, art, and geography lessons to complete the intervention. However, most respondents stated that they rotated the times at which pupils were taken out of class each week so as to avoid them missing the same lessons and falling behind in a particular subject (42 out of 63 respondents).

Some staff felt that by holding (most) of the sessions in the afternoon, they did not always get the best out of pupils as they can be tired and their concentration is generally lower than in the mornings. In addition, through the session observations, it was clear that pupils could be reluctant to engage with the intervention if they had been taken out of an afternoon lesson they enjoy (e.g., PE). However, core subject teaching generally took priority, and some schools did not have access to the necessary resources (i.e. quiet teaching space and computers with internet access) until the afternoon in any case, which gave them little flexibility in terms of when sessions could be delivered.

In terms of how easy or difficult treatment schools found the process of securing two computers and a quiet teaching space to deliver the intervention, the experiences of implementers were again mixed. Broadly, this was dependent on how well resourced the school already was and/or the amount of forward planning that had been undertaken to ensure that these resources were available to TAs over the trial period.

With regards to accessing two computers, for instance, 18 out of 38 TAs that responded to this question in the end of project survey stated that they found the process easy (with 12 respondents stating that they found it very easy), while 13 had found it difficult. When asked to explain their response, as indicated, those who found the process easy either already had plenty of computers available within the school to meet current demand and therefore enough to spare for delivering the intervention, and/or had secured two laptops for the duration of the trial that were solely used by implementers for this purpose. Those that had taken the latter approach spoke of how they had noted the importance of having these resources available to the TA after the training day, and had taken steps to source these when they returned to the school. During the fieldwork visits, one school mentioned that they had used the small bursary they received for joining the project to purchase two new laptops that could be used exclusively for the trial.

Among those who found the process of securing two computers difficult, it appeared from the survey results and fieldwork visits that these TAs had to secure the computers needed to deliver the sessions on an ad hoc basis (either weekly or daily). In schools where computers were in short supply or were old (and therefore slow and unreliable), this created additional difficulties in trying to find the appropriate resources necessary to complete the sessions and in keeping to time.

The experience of implementers in securing a quiet teaching space for the intervention was similar. Eighteen out of 46 TAs had found the process of sourcing a quiet room easy, while 16 had found it difficult. Among those who found the process easy, or offered a neutral response (12 respondents), they commented that they regularly had access to a room that was dedicated to the delivery of pupil interventions, or other teaching spaces within the school were generally available at the times required (though a few reported that these spaces were double booked on occasion). Several respondents further added that they recognised that they were fortunate to have this dedicated teaching space in a busy school environment.

In contrast, those that found the process difficult observed that suitable teaching spaces were in very short supply within their school. Some reported that lots of pupil interventions were taking place at the same time, so quiet spaces were at a premium and they occasionally had to share rooms with TAs and pupils undertaking other quiet interventions. Where suitable teaching spaces were not available, a few respondents recalled having to use staff offices, the school staff room and/or open spaces between classrooms to deliver the sessions. As already noted, where sessions were not delivered in quiet teaching spaces due to a lack of availability (e.g., in corridors), this affected pupils' ability to complete the independent online games. Further, having to search for suitable spaces on an ad hoc basis, as some schools did, shortened the time the TA had available to deliver the session.

4.8 Implementation of support system

The project included a few features that were meant to support high-quality implementation. These included establishing a link teacher to support the TA, and on-going support from the delivery team over the trial period. The formal role of the link teacher, as specified by the delivery team, was to plan an agreed timetable for the programme with the TA; assist in securing the resources necessary to deliver the intervention (two computers and a quiet teaching space); act as the TA's mentor; support the delivery of the programme and ensure sufficient preparation time; and oversee any communication with parents. These responsibilities would sometimes require liaising with the year 3 teacher(s), if the link teacher did not hold this post.

It appeared from the fieldwork visits that link teachers had generally fulfilled these responsibilities, with TAs saying that they had received good support throughout the project. A few link teachers observed that their attendance at the training day(s) had helped greatly in this regard. In particular, link teachers spoke of being able to be the first point of contact for the TA if they were unclear about how to deliver any of the activities, and of being able to go through the handbook with them given that they had some understanding of the project.

'There are times that [the TA] *talks to me to try and troubleshoot or, 'I'm not quite getting this', so we'll look through the handbook together and we'll figure it out. So there are times when I can do that. And I guess in that sense, the training has its benefit, because otherwise I'd be really starting from scratch'.*

Link teacher, School 3, WM+ group

Staff in treatment schools were also very positive about the on-going support that they received from the delivery team after the initial training session and starting implementation. Almost all schools visited commented that communication with the team had been very good, and that they were always quick to respond to any queries that they had via email. The delivery team also sent round the responses to FAQs to all schools, which were perceived to be helpful, and in some cases provided the answer to issues that some schools experienced later in delivery.

Several TAs commented that the interim visits conducted by the delivery team half-way through implementation - where they observed TAs delivering sessions and provided support and feedback – were particularly useful in clarifying any uncertainties they had about how to interpret the handbook instructions, or in providing reassurance that they were delivering the exercises correctly.

4.9 Outcomes

As noted, the process evaluation explored the perceived outcomes of the intervention among those overseeing implementation. While staff in treatment schools were willing to put forward their views, it should be noted that several limitations were highlighted that affected the certainty with which they could say that the intervention had been effective for participants.

Firstly, in most cases, the fieldwork visits took place before implementation had finished. As such, pupils had yet to receive the complete trial 'dosage' and its full impact was therefore still unclear. Second, as the delivery of the sessions took up a significant amount the TAs' working week, especially among those that worked part-time, they were unable to observe whether pupils were transferring the rehearsal strategies and/or knowledge of arithmetic concepts they had practiced in the sessions to the classroom. This was also affected by pupils in larger schools sometimes being spread over two to three year 3 classes. Some link teachers were similarly unable to comment on whether pupils were transferring these strategies or knowledge to the classroom as they were unaware of everything that had been covered in the sessions or were not the year 3 class teacher. Finally, even where progress among pupils was noted, some respondents found this difficult to disentangle from pupils' natural development at this age, as well as the impact of other formal interventions and additional support that they had been receiving.

With these caveats in mind, some treatment schools did tentatively identify improvements in a few pupils' abilities and performance, as well as their attitude and approach to learning in class. For instance, as Table 18.3 shows, while 28 percent of respondents could not say what the impact on pupils had been (or saw no impact) around a quarter of respondents (24%) stated that they had seen improvements in pupils' ability to retain information.

Table 18.3 - Among those pupils that were involved in the programme, have you seen/heard of noticeable changes in any of the following? (select all the apply)

	Overall	WM	WM+
Their attention and behaviour in class	19 (16%)	11	8
Their ability to retain information	29 (24%)	15	14
Their general performance in English	9 (8%)	7	2
Their general performance in maths	30 (25%)	11	19
Don't know/Can't tell/None of the above	33 (28%)	20	13
Total number of responses	120 (100%)	64	56

The interview data also explored the perception among TAs and link teachers of changes in pupils' abilities and performances. While staff stressed that it was difficult to isolate the impact of the programme, several link teachers had noticed an increased alertness and focus among one or two pupils participating in the programme in the year 3 class since the beginning of the trial. They felt that these pupils were now actively listening to and retaining teachers' instructions and had a clearer idea of the aims of each lesson. This included some pupils who had speech and language difficulties, or in another case ADHD. Some link teachers also stated that they had observed pupils using the rehearsal strategies unprompted to help them recall key bits of information in class. In most cases, teachers saw pupils applying these strategies during spelling tests to arrive at the correct answer. They commented that this showed that pupils were happy to adopt the rehearsal strategies when they could see them working for them.

Another change commonly identified by treatment schools was an improvement in pupils' general performance in maths. Again, a quarter of respondents to the end of project survey had noticed these changes among participants. Across treatment groups, the number of respondents in the WM+ group who had seen some improvement was somewhat higher. This is most likely indicative of the time spent completing arithmetic exercises in this arm of the trial, though the way in which pupils benefitted from and transferred this experience to the classroom was not clear from the fieldwork visits. They showed that schools' approaches to teaching mathematics clearly differed, and link teachers had varying views on how well the arithmetic games complemented the topics covered in maths lessons. When teachers spoke of having seen improvements in mathematics, this included schools from the WM group, who again noted examples of pupils independently applying the rehearsal strategies to recall multiplication tables or to conduct calculations with multiple steps in their heads.

'One pupil has shown an increased understanding in maths and she's a lot more able to do, attempt problems with multiple steps, so finding fractions of amounts for example, I don't whether the rehearsal strategies are coming into it in her head, but she going okay, total amount divided by the denominator times by the numerator, so you've seen that process and she's been able to use that independently. So with some of them you can see the strategies transferring over'

Link Teacher, WM group, School 8

Other improvements noted by respondents that were not included in the list of survey responses were increased confidence among pupils in class. This was also noted among several staff members during the fieldwork visits. A few interviewees speculated that the regular success that pupils had in progressing through the TA-led activities and independent games, for which they were provided with individual praise and encouragement, may have had a positive motivational effect on pupils and increased their confidence in their abilities. As shown, learning the rehearsal strategies had provided some pupils with an additional tool with which approach problems in class. Other staff therefore felt that this increased confidence stemmed from pupils being able to respond to challenges with a greater degree of calmness and certainty that they could arrive at the correct answer with the aid of these techniques.

Respondents were also asked among how many pupils they had observed these perceived changes. While 17 out of 76 respondents to this question had not noticed any of these changes in the pupils who had taken part, 25 noted that they had seen these improvements in around half of pupils, while 21 respondents had seen them in a minority. As Table 18.4 shows, there was some differentiation between treatment groups here, with respondents from the WM group perceiving these changes to be slightly more widespread. Again, the exact reason for this difference is unclear, though it may reflect the comments from some staff in the WM+ group that a few pupils found the arithmetic games quite challenging and in a few cases inaccessible.

Table 18.4 – Among how many pupils have you seen the changes noted?

	Overall	WM	WM+
None of the pupils	17	8	9
A minority of pupils	21	6	15
Around half of pupils	25	15	10
A majority of pupils	8	6	2
All pupils	5	4	1
N	73	39	34

The fieldwork visits explored differences in impact further. A consistent view expressed across some treatment schools was that the intervention had no impact on those pupils who had pre-existing behavioural problems. They observed that this group of pupils found it difficult to engage with the intervention and concentrate for the required period, especially when completing the independent games. A few staff members did note, however, that the project had been helpful in highlighting what additional support these pupils required before they could be ready to engage in an intervention such as this.

4.10 Other interventions and support pupils had been receiving

As highlighted, it was clear from both the end of project survey and fieldwork visits that some pupils who participated in this intervention were also receiving additional support with their learning, which may also explain any progress they had made over the course of the academic year. Most commonly, this took the form of additional support in maths. For instance, 30 respondents out of 86 noted that some pupils participating in the programme had received other interventions in this area. In almost all cases, it appeared that this additional support was delivered on a weekly basis in single or multiple sessions, and could include up to half of participating pupils in the WM programme. This took the form of either formal programmes (e.g. First Class at Number, Numicon, Power of 2), or general maths ‘booster’ groups with a select number of pupils who struggle with the subject to consolidate what they had learnt in class. Where specified, respondents stated that this additional support aimed to assist pupils with their basic maths skills such as their mental maths ability, times table recall, understanding of number bonds and place value.

Survey respondents were also asked whether participating pupils had received any additional support with their memory over the course of the academic year. Most (70 out of 85 respondents) reported that pupils had not received any other interventions in this area. In the few instances where pupils had received additional support, respondents mentioned that some pupils had participated in formal programmes (e.g. Memory Matters, Auditory Memory, and Memory Fix), though it was not clear how regularly these sessions took place.

4.11 Negative effects

Through the fieldwork visits, the process evaluation also explored whether participating in the intervention had any unintended consequences or negative effects on treatment schools. The most common comment from staff was that the intervention was very time intensive. As a result, link teachers remarked that the TAs delivering the sessions were taken away from the classes/pupils they typically support for a large amount time each week. Where implementers were the year 3 class TA, some teachers from this year group stated that they had to fill in for the TA in their absence and spend a greater amount of time than they would usually assisting those children with additional support needs. In their view, this had a detrimental impact on the rest of the class who lost some of their teaching time as a result. In other cases, year 3 teachers had to source cover for the TA, which put additional demands on school resources.

Where the TA implementing the programme was dedicated to delivering interventions these issues could be avoided, though one school did comment that the time commitment the programme required had prevented the intervention TA from providing support to other year groups.

In a few instances, staff from treatment schools were not fully aware of the extent of the time commitment that the project would require from TAs when they first signed up. This appeared to be due to a lack of understanding or poor communication from the senior leader within the school that committed to the programme. As such, once the delivery requirements became clear, in these instances, schools had to increase the TAs working hours for the duration of the project to ensure that they had enough time to deliver all the sessions in addition to their other commitments. This presented an additional and unexpected financial cost to the schools concerned.

4.12 Formative findings

Based on treatment schools' experience of implementing the intervention, the process evaluation explored their views on how the WM programme could be improved to help inform its future rollout. The suggested alterations generally centred on two aspects: the independent online games and resourcing the intervention within the school. As the preceding discussion has shown, these were the areas where treatment schools encountered the most difficulties during implementation.

With regards to the independent games, as noted, some pupils had found it difficult to complete these and not lose focus over the course of 30 minutes. As such, a few respondents to the end of project survey suggested that if they were to deliver the programme again, they would either just carry out the TA-led activities, in which pupils generally progressed well with the TAs support, or they would deliver the sessions with one pupil at a time as opposed to two, so they could be on hand to support the child with the online game should they encounter any problems. Others stated that they would prefer all participants to complete the independent games in a separate group session where the TA could be on hand to provide prompts to pupils and support where needed, and not have their attention taken away by the TA-led activities.

As mentioned previously, another difficulty that treatment schools had in implementing the programme was the amount of time TAs had to commit to delivering the sessions in their working week, and the additional pressures this put on school resources as a result.[7] Staff in treatment schools thereby suggested several ways in which to reduce the pressure placed on school resources by the project and the time intensity of the intervention for individual TAs. One school stated that if they were to run the intervention again, they would select a smaller number of pupils to participate in the programme under

[7] Details of how sessions were delivered were only collected from TAs

the minimum threshold of 10 set by the delivery team. They commented that this model would be more sustainable in terms of the resources required and may also allow them to deliver the sessions over a longer period in order to increase their potential impact. Another school indicated that they would have two TAs delivering the sessions, instead of one (NB some treatment schools did adopt this model). This would help ensure teachers were not without their class TA for prolonged periods. A final suggestion was for the programme to be shorter in duration, perhaps carried out over the course of a term, to take up fewer school resources (e.g. TA time, IT facilities) over the course of the academic year.

Other separate suggestions for how the programme could be improved that were put forward by individual respondents included:

- Further guidance for year 3 teacher in how to build on and enhance the skills and strategies learnt in the intervention within class
- A video example of a session being delivered by a TA as part of the formal training
- Further training time for arithmetic games (i.e. a full day as opposed to half a day given that there are more games for TAs to learn the rules of compared to the working memory games)

4.13 Control group activity

An online survey was administered to schools allocated to the control group at the end of the 2016/17 academic year. Its purpose was to gather information on what activity had been undertaken with year 3 pupils who were selected to participate in the trial prior to randomisation. Specifically, schools were asked what additional support selected year 3 pupils had received over the trial period to support the development of their number skills and recall ability in the absence of the Improving Working Memory programme. These findings may help to explain any impact (or lack of impact) observed between schools allocated to the treatment and control groups. However, it should be noted that the survey was completed by just over half of schools allocated to the control group (24 out of 43) and so only provides a partial insight into their activities over this period.

Similar to treatment schools, there was a substantial minority of control schools where pupils had received additional support in maths: 11 out of 26 respondents stated that year 3 pupils participating in the trial had received other interventions focused on improving their performance over the past year, while 14 reported that pupils had not received anything of this nature. As in the treatment group, this additional support commonly took the form of several weekly sessions delivered either in small groups or on a one-to-one basis depending on pupils' level of need. A few respondents commented that these sessions were targeted at gaps identified in pupils' basic maths skills, while others referred to particular programmes or methods that were being used to address these needs. The programmes cited included: Num skills, Power of 1/Power of 2, and Number Sense; a few respondents also mentioned that they were using precision teaching methods to assist pupils, for instance, to learn their times tables more quickly. These sessions lasted from 20-60 minutes depending on their frequency (e.g. sessions that were delivered several times a week were reported to be shorter in duration).

Control schools were also asked whether year 3 pupils selected to take part in the trial had received any interventions focused on improving their memory over the past academic year. As with treatment schools, few reported that pupils had received any additional support in this area (3 out of 26 respondents). Where it occurred, this additional support again took the form of short weekly sessions where pupils received precision teaching to help retain particular language skills (e.g. phonics, spelling etc.)

Conclusion

Key conclusions
1. Children in both the WM and WM+ schools made the equivalent of 3 additional months' progress in maths, on average, compared to children in the business as usual control schools. These results have high security ratings.
2. Pupils eligible for Free School Meals (FSM) in the WM schools made a small amount of additional progress in maths compared to those in the control schools and no impact was found for FSM children in the WM+ schools. These results are of lower security than the overall findings because of the smaller number of pupils.
3. The evaluation found positive impacts on working memory, and attention and behaviour in class for pupils receiving the interventions compared to children in comparison schools.
4. The intervention was found to be time intensive, predominantly due to the need for TAs to leave class to deliver sessions, which increased pressure on teachers during lessons and in some cases required schools to source TA cover.

5.1 Interpretation

In line with existing evidence for both one-to-one tuition and individualised learning, the current trial demonstrates that this combination of support can deliver positive effects for learners. The main result of the trial was that the Improving Working Memory Programme had a positive impact on recall, and attention and behaviour in class.

The impact on attainment is less clear. While the estimated impact on attainment was positive, we cannot attribute this to the intervention with much certainty given this effect falls short of statistical significance at the conventional level. However, we note also that the size of the estimated effect on attainment has turned out to be smaller than the trial is powered to detect. In view of this, we emphasise that the appropriate interpretation is not that the intervention necessarily had no effect on this outcome, but that it did not have a sufficiently large effect for the trial to detect.

On the other hand, the secondary outcomes do provide evidence of significant impact. The programme resulted in improvements in recall across a number of dimensions. These are the outcomes that the programme is designed to influence most directly, and it seems reasonable to speculate that the improved recall ability might eventually translate into attainment gains. In a similar vein, the positive impacts on attention and behaviour may have longer-term benefits.

There were substantial limitations that affected the certainty with which schools could say the intervention had been effective for pupils. The process evaluation revealed that in many cases, TAs spent limited time in classrooms during the intervention and as a result were unable to identify whether strategies were used by pupils outside of the working memory sessions. These findings are based on a survey of treatment schools with a 75% response rate, as well as 8 fieldwork visits to providers across the WM and WM+ groups.

The Education Endowment Foundation's toolkit summary on individualised instruction states that while this tends to have a positive effect on learners, the average impact is low, with some studies reflecting small negative impacts for learners. Both the WM and WM+ interventions included an element of individualised learning, in combination with one-to-one tuition, providing interesting indicative evidence for the efficacy of these mechanisms in combination.

An important element of personalised learning is that pace is determined by a pupil's individual level of ability. Pupils were supported with games by TAs, in which their progress through the levels of the game was determined by their performance on the previous level. While games increased in difficulty as pupils progressed through levels, analysis of the game data appears to suggest pupils' performance improved as games were repeated.

In addition, pupils in each treatment arm were provided with one-to-one support. The EEF toolkit also suggests that where such support is provided as a replacement for other lessons (as it was during the current trial) it accelerates learning on average by approximately five additional months' progress. It is interesting to note that of these studies, optimum impact is associated with a much higher dosage (30 mins, 3 - 5 times a week over a 12-week period) than what is delivered in the current trial. Given this, it is reasonable to suppose that with increased dosage over the 10-week delivery period, larger effects may have been observed.

Though improvements in working memory were not reflected in attainment outcomes in the current trial, we note that as an efficacy trial, impacts are estimated under the most favourable conditions. It may be that an effectiveness trial with its larger sample size would increase the chances of finding an effect of this size to be significant. Offsetting this is the possibility that under less favourable conditions the effect would be smaller. The impact on working memory outcomes is however, promising, and indicate the importance of cultivating domain general abilities such as working memory alongside more specific abilities like numeracy skills to help prevent future learning difficulties.

5.2 Limitations

Whilst measurement attrition for endline math attainment was relatively low, the imbalance in KS1 pretest data reduced the security of the findings. Where previously KS1 data could have been obtained from the National Pupil Database (NPD) for schools who were not able to produce it, point scores are no longer recorded for the period in which participating pupils sat their KS1 exams.

Balance checks revealed some imbalance across trial arms, with on average lower Reasoning scores in the control group in comparison with both treatment arms, and the reverse pattern for Arithmetic scores. However, it is reassuring that this did not result in statistically significant differences across treatment arms in the KS1 point estimates. In addition, the imbalance across arms on the proportion of pupils ever eligible for FSM may have had a small impact on the final reported effect size. However, robustness checks including FSM status as an additional covariate did not affect the point estimates.

Following randomisation five schools advised they were unable to commit to the delivery of the programme, and whilst it was still possible to collect post-test data for these pupils, the schools were deemed non-compliant. Four of these schools had been allocated to the WM+ arm of the trial, and thus the loss of compliance may have diluted the final result for this arm. Though the EEF cost rating for the programme (based on financial costs) has been determined to be 'very low', some schools, including those unable to commit to delivery, reported that the intervention itself was or would be time intensive, creating significant additional demands on school resources. As a result, some schools suggested that were the intervention to be rolled out more widely, a smaller group of pupils or a shorter programme would reduce the impact on staff time.

Generally, as discussed earlier, the data collected from treatment schools as part of the process evaluation (either via survey or fieldwork visits) only represent the views and experiences of a subset of the larger treatment population. Whilst visited schools were selected to include a variety of delivery contexts, the qualitative findings are not necessarily representative, and should be considered within the context of its limitations. There may be some recall errors in survey responses in terms of estimating how many pupils started and completed the intervention, and how many SEN pupils participated. The uncertainty of the latter was clear from the fact that estimates differed between survey respondents within the same school, possibly resulting from recall error or different understandings/interpretations of SEN. The majority of schools were later found to have been compliant with guidelines provided, with a small number of statemented children (three) participating in the trial.

5.3 Future research and publications

Given the current indicative evidence for the positive impact of the intervention on Math attainment, and the relatively low cost of delivery, future trials might look to explore variations of the intervention in which it is delivered in small groups, which EEF evidence suggests yields similar, and in some cases larger effects. The quality of the TA training, and the balance between time spent receiving on- to-one support and time spent on practice are other elements for which variation could be explored. It may also be worthwhile to examine the effect of increased dosage, in which pupils are offered more one-to-one tuition with TAs.

In each case, it would be important to increase fidelity to the intervention design where possible, specifically to ensure the selection of pupils for participation takes place as per the developers' instructions, and to measure and manage the division of time spent by TAs across the one-to-one support and practice sessions.

It is the intention of the developer, Oxford University, to publish the current evaluation.

References

Alloway, T.P., 2007. *Automated Working: Memory Assessment: Manual*. Pearson.

Alloway, T.P., Gathercole, S.E., Kirkwood, H. and Elliott, J., 2009. The cognitive and behavioral characteristics of children with low working memory. *Child development*, *80*(2), pp.606-621.

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

Baddeley, A.D., Allen, R.J. and Hitch, G.J., 2011. Binding in visual working memory: The role of the episodic buffer. *Neuropsychologia*, *49*(6), pp.1393-1400.

Ebbes, P. Bockenholt, U. and Wedel, M. (2004) Regressor and random-effects dependencies in multilevel models. *Statistica Neerlandica, 58*(2), pp.161-178.

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

Holmes, J., Gathercole, S.E. and Dunning, D.L., 2009. Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental science*, *12*(4).

Holmes, J. and Gathercole, S.E., 2014. Taking working memory training from the laboratory into schools. *Educational Psychology*, *34*(4), pp.440-450.

Nunes, T., Evans, D., Bell, D. and Campos, T., 2008. Improving children's working memory through guided rehearsal. *American Educational Research Association (AERA), New York: April*.

Nunes, T., Barros, R., Evans, D. and Burman, D., 2014. Improving deaf children's working memory through training. *International Journal of Speech & Language Pathology and Audiology*, *2*(2), pp.51-66.

Passolunghi, M.C. and Costa, H.M., 2016. Working memory and early numeracy training in preschool children. *Child Neuropsychology*, *22*(1), pp.81-98.

Swanson, J.M., Kraemer, H.C., Hinshaw, S.P., Arnold, L.E., Conners, C.K., Abikoff, H.B., Clevenger, W., Davies, M., Elliott, G.R., Greenhill, L.L. and Hechtman, L., 2001. Clinical relevance of the primary findings of the MTA: success rates based on severity of ADHD and ODD symptoms at the end of treatment. *Journal of the American Academy of Child & Adolescent Psychiatry*, *40*(2), pp.168-179.

Worth, J., Sizmur, J., Ager, R. and Styles, B., 2015. Improving numeracy and literacy. *London. https://goo.gl/9TcCLz*.

Appendix A: EEF cost rating

Cost ratings are based on the approximate cost per pupil per year of implementing the intervention over three years. More information about the EEF's approach to cost evaluation can be found here. Cost ratings are awarded as follows:

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

Appendix B: Security classification of trial findings

Rating	Criteria for rating			Initial score	Adjust	Final score
	Design	Power	Attrition[8]			
5	Well conducted experimental design with appropriate analysis	MDES < 0.2	0-10%	5		
4	Fair and clear quasi-experimental design for comparison (e.g. RDD) with appropriate analysis, or experimental design with minor concerns about validity	MDES < 0.3	11-20%		Adjustment for Balance **[-1]**	4
3	Well-matched comparison (using propensity score matching, or similar) or experimental design with moderate concerns about validity	MDES < 0.4	21-30%			
2	Weakly matched comparison or experimental design with major flaws	MDES < 0.5	31-40%		Adjustment for threats to internal validity **[0]**	
1	Comparison group with poor or no matching (E.g. volunteer versus others)	MDES < 0.6	41-50%			
0	No comparator	MDES > 0.6	over 50%			

- **Initial padlock score:** lowest of the three ratings for design, power and attrition = [5] padlocks. Well conducted experimental design with a MDES at randomisation of 0.19. Outcome attrition was low at 8.3%, but it has to be noted that missing pre-tests were imputed.
- **Reason for adjustment for balance** (if made): [-1] One padlock was dropped due to relevant imbalances in school characteristics and the number of FSM pupils. Sensibility analyses suggested by reviewers indicated that these imbalances did not affect the main results. There were also imbalances in the pre-test data (KS1, Arithmetic and Reasoning) but they were always below 0.05.
- **Reason for adjustment for threats to validity** (if made): [None identified]
- **Final padlock score:** initial score adjusted for balance and internal validity = [4] padlocks

[8] Attrition should be measured at the pupil level (even for clustered trials) and from the point of randomisation to the point of analysis.

Appendix C: Randomisation code

```
set seed 3636378

*create blocks
xtile   att3_all  = newpcexpectedorhigher, nq(3)
gen     block_all = .
      replace block_all = 1 if twoplusformentry==0 & att3_all == 1
replace block_all = 2 if twoplusformentry==0 & att3_all == 2
replace block_all = 3 if twoplusformentry==0 & att3_all == 3
replace block_all = 4 if twoplusformentry==1 & att3_all == 1
replace block_all = 5 if twoplusformentry==1 & att3_all == 2
replace block_all = 6 if twoplusformentry==1 & att3_all == 3
// Where unknown, to mean for 1 or 2+ form entry, as appropriate
replace block_all = 2 if twoplusformentry==0 & block_all==.
      replace block_all = 5 if twoplusformentry==1 & block_all==.

#delimit ;
lab def block
1 "Oneform_upperthird"
2 "Oneform_middlethird"
3 "Oneform_lowerthird"
4 "Twoform_upperthird"
5 "Twoform_middlethird"
6 "Twoform_lowerthird";
#delimit cr
lab val block_all block
lab var block_all "Randomisation block, overall"

***Randomisation
* randomly assign first school, subsequent schools assigned sequentially
gen double rand=uniform()
sort block_all rand
gen T=irecode(rand,.33333,.66667,1)
local i=1
while `i'<_N {
    local i=`i'+1
    replace T=0 if T[_n-1]==2
    replace T=1 if T[_n-1]==0
    replace T=2 if T[_n-1]==1
}
lab def T 0 "Control" 1 "WM" 2 "WM+"
lab val T T
lab var T "Treated"
```

Appendix D: Memorandum of Understanding

UNIVERSITY OF OXFORD
DEPARTMENT OF EDUCATION

Professor Terezinha Nunes
15 Norham Gardens, Oxford OX2 6PY
Tel: +44(0)1865 284893
www.education.ox.ac.uk

Working Memory and Number Skills in Year 3

Information sheet

The Education Endowment Foundation (EEF) is supporting a project to evaluate a programme designed for Year 3 pupils who are struggling with number skills. Children with low number skills often have limited working memory, which is the ability to keep information in mind while at the same time working on the information. Our working memory programme was found effective in improving children's working memory capacity in two previous studies. In this project, we aim to evaluate whether improvements in working memory have an impact on number skills.

Because this is a research project, schools will be randomly assigned to implement the working memory programme either in 2016 or in 2017. The research team from Oxford University will train a teaching assistant and a lead teacher in each participating school to use the programme. The TA will then be able to work with pupils who are struggling with number skills at the beginning of Year 3. The minimum number of pupils participating is 10 per school but a school with 2/3/4 form entry can identify up to 20 pupils. Schools will be free to continue using the programme with other pupils who might benefit from it after the end of the research project.

Participating schools will receive training free of charge, all the materials, and a fixed budget of £725 to cover the TA and the lead teacher replacement for the day spent in training. Schools will need to provide a computer or tablet with Internet access for the TA to use when working with pupils.

In line with other work supported by the EEF, the evaluation of the programme will be carried out by randomly assigning schools to participate in the working memory intervention in 2016 or to a waiting-list group. The waiting-list group will receive the same fixed budget as the group participating in the training in 2016 and will have the choice of using it to participate in equivalent training in 2017 or to use the amount for another training of its choice. All schools in the project will receive their budget after the data collection has been completed in July 2017.

The project involves a partnership between the University of Oxford, NIESR (National Institute of Economic and Social Research, BIT (Behavioural Insights Team) and the Education Endowment Foundation (EEF), which is providing the resources for the project. The teams from NIESR and BIT are responsible for the independent evaluation of the working memory training and its impact on number skills.

Please keep this information sheet for your records.

Memorandum of Understanding regarding the project Improving Working Memory

This agreement is between the School named below and the Department of Education, University of Oxford, about a randomised control trial of the intervention to improve children's working memory, funded by the Education Endowment Foundation and evaluated by a team from NIESR and BIT.

Head Teacher:	
Telephone:	e-mail:

The School

- The School understands that it will be randomly allocated either to the group that participates in the working memory programme in 2016 or to the group that has the option to use its budget to participate in the programme in 2017; it commits to full participation in either group and to post-testing between May and July 2017.
- The School will identify a group of Year 3 pupils who are struggling with number skills in early September 2016; 10-20 pupils (depending on the size of the cohort) will be nominated as participants to the Oxford University project team.
- The School will seek permission from the parents of pupils participating in the programme for the working memory and number skills assessment and for the data to be shared with the evaluation team; separate permission for sharing the pupil's unique pupil identifier will be sought in the same consent form.
- In September/early October 2016, the School will provide access to the research teams to administer a working memory assessment to the participating pupils and provide completed classroom behaviour ratings by the teacher of the participating pupils; the rating scale is provided by Oxford University and contains 15 simple questions.
- If the School is allocated to the working memory programme in 2016, the school will nominate a teaching assistant and a link teacher to attend the training session and to manage and deliver the intervention in accordance with guidance from OU; for successful implementation, the TA needs access to a computer/tablet with internet access, time to work individually with each pupil, and to print and distribute the pupils' certificates of completed activities, provided electronically by OU.
- The School will communicate fully and promptly with OU and the evaluation team, share appropriate data and ensure that questionnaires and surveys are completed and returned.
- The School will facilitate visits to the school by OU to support the implementation of the intervention and by the evaluation team to observe and interview staff during 2016-17 and to administer a post-test to participating pupils between May and July 2017.
- The School understands that it will receive intervention training, resources and support free of charge in return for complete participation in the trial as set out in this agreement.

Oxford University

- OU will provide an intervention training programme for the School's teaching assistant and link teacher:
  - if the evaluation team allocates the School to the working memory training in 2016, training will be in late November/early December 2016;
  - if the evaluation team allocates the School to the 2017 group, training will be in June/early July 2017, after all pupils have been post-tested.
- OU will inform the School of the venue and dates of its training programme when the evaluation team has made the allocation, endeavouring to allocate the School to its first preference wherever possible.
- OU will provide guidance and all the materials required for the delivery of the working memory intervention.
- OU will provide a support visit to the School and certificates for pupils to reward their performance, which the school can print and share with the pupils.
- OU will provide e-mail and telephone helplines for schools during the life of the project (until March 2018).
- OU will promptly transfer the budget for TA and teacher replacement to the school as soon as the final assessments are carried out and all the data have been provided to the evaluation team.

Signed for and on behalf of:

Roberttown CE J&I School	Oxford University
Signed: ______________________	**Signed:** ______________________
Name: *(print)* ______________________	**Name:** Terezinha Nunes
Position: Headteacher	
Date: ______________________	**Date:** ______________________

Please return 1 signed copy of this agreement using the envelope provided and keep a copy for your records.

Appendix E: Supplementary analysis

Sensitivity of the results to the imputation approach

The results presented in Table 12 used multiple imputation to address the problem of missing values among the regressors. Table D1 shows the results when using an alternative approach of null imputation, or the alternative, without imputation.

Table D1: Impact estimates using null imputation and no imputation

	Raw means						Effect size		
	WM group		WM+ group		Control group			WM group	WM+ group
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (WM; WM+ control)	Hedges g (95% CI) p-value	Hedges g (95% CI) p-value
Primary outcome									
Null imputation	481 (38)	8.8 (3.3, 14.3)	482 (43)	8.8 (3.3, 14.3)	512 (42)	8.5 (2.6, 14.4)	1352 (443; 439; 470)	0.16 (-0.13, 0.45) 0.29	0.26 (-0.01, 0.53) 0.06
No imputation	481 (38)	8.8 (3.3, 14.3)	482 (43)	8.8 (3.3, 14.3)	512 (42)	8.5 (2.6, 14.4)	1274 (426; 397; 451)	0.18 (-0.10, 0.46) 0.21	0.27 (0.02, 0.52) 0.03

Analysis with pooled treatment arms

The results of supplementary analysis in Table D2, in which WM and WM+ are pooled, reflect the findings of the original analysis. The effects for both arms are not significant.

Table D2: Pooled Treatment Arms

	Raw means						Effect size		
	WM/WM+ group				Control group			WM/WM + group	
Outcome	n (missing)	Mean (95% CI)			n (missing)	Mean (95% CI)	n in model (WM/WM+ control)	Hedges g (95% CI) p-value	
Primary outcome									
BAS3 score	963 (81)	8.8 (3.3, 14.3)			512 (42)	8.5 (2.6, 14.4)	1352 (882; .; 470)	0.21 -0.03, 0.45) 0.09	
Secondary outcome									
Listening Recall	963 (43)	10.6 (4.3, 16.9)			512 (17)	9.9 (4.3, 15.5)	1415 (920; .; 495)	0.30 (0.10, 0.50) 0.00	
Counting Recall	963 (43)	17.1 (8.5, 25.6)			512 (17)	15.6 (7.7, 23.4)	1415 (920; .; 495)	0.68 (0.40, 0.96) 0.00	
Backward Digit Recall	963 (43)	11.0 (3.7,			512 (17)	10.2 (4.0,	1415 (920; .;	0.41 (0.18,	

		18.3)				16.4)	495)	0.65) 0.00	
WM factor	963 (43)	0.1 (-1.9, 2.1)			512 (17)	-0.2 (-2.0, 1.5)	1415 (920; .; 495)	0.29 (0.16, 0.42) 0.00	
Attention and Behaviour in Class	963 (49)	28.4 (8.2, 48.6)			512 (29)	29.3 (8.8, 49.9)	1397 (914; .; 483)	-0.57 (-1.02, -0.13) 0.01	

Analysis controlling for FSM status

As the imbalance across arms on the proportion of pupils ever eligible for FSM may have had a small impact on the final reported effect size, a final version of the analysis was conducted to control for FSM in the model. The results are presented in table D3 below. The effects for both arms remain non-significant.

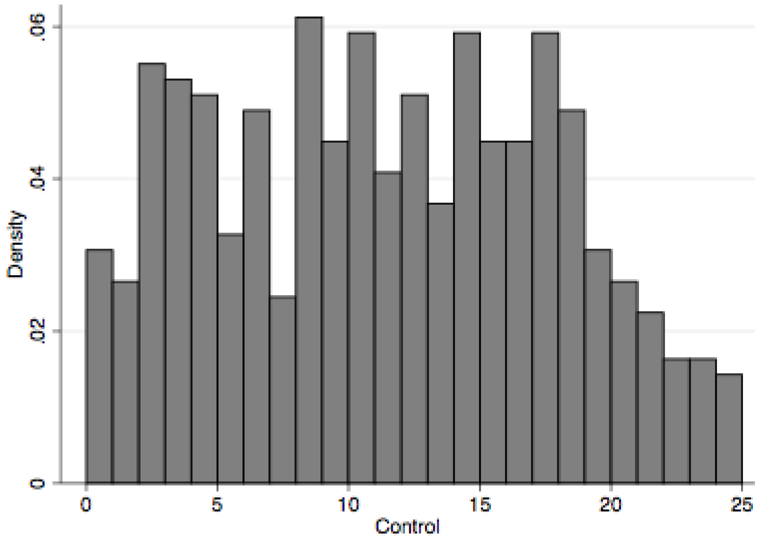
Table D3: Analysis controlling for FSM status

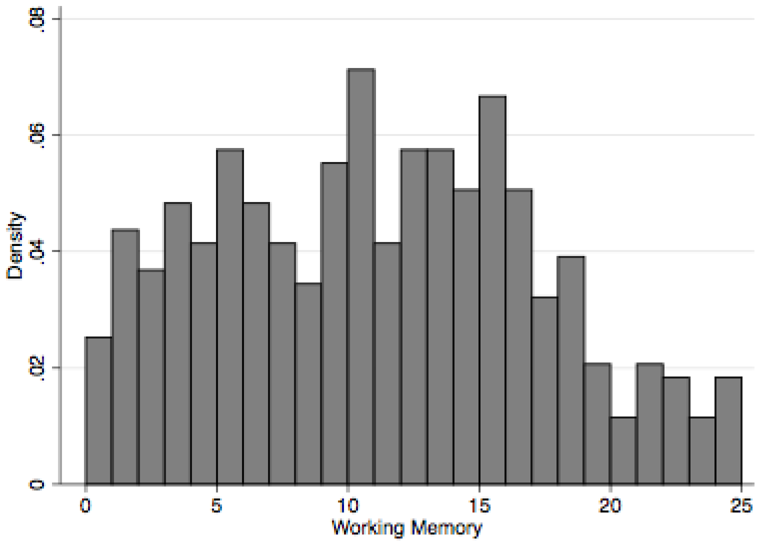
	Raw means						Effect size		
	WM group		WM+ group		Control group			WM group	WM+ group
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (WM; WM+ control)	Hedges g (95% CI) p-value	Hedges g (95% CI) p-value
Without FSM regressor									
BAS3	420 (30)	8.8 (3.2, 14.3)	380 (34)	8.9 (3.3, 14.5)	444 (33)	8.6 (2.9, 14.3)	1147 (390; 346; 411)	0.08 (-0.22, 0.37) 0.62	0.20 (-0.12, 0.51) 0.22

With FSM regressor									
BAS3	420 (30)	8.8 (3.2, 14.3)	380 (34)	8.9 (3.3, 14.5)	444 (33)	8.6 (2.9, 14.3)	1147 (390; 346; 411)	0.09 (-0.20, 0.38) 0.56	0.19 (-0.11, 0.48) 0.21

Appendix F: Histograms

Figure 6: Histograms of KS1 Arithmetic Scores by Group





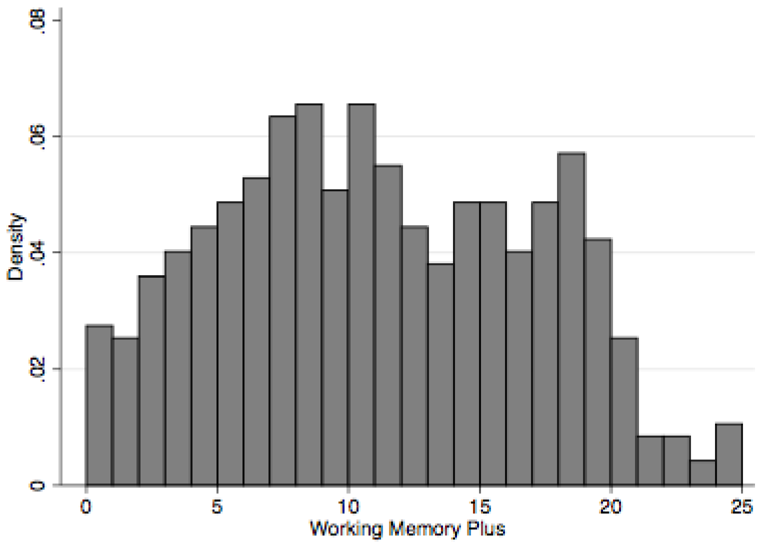
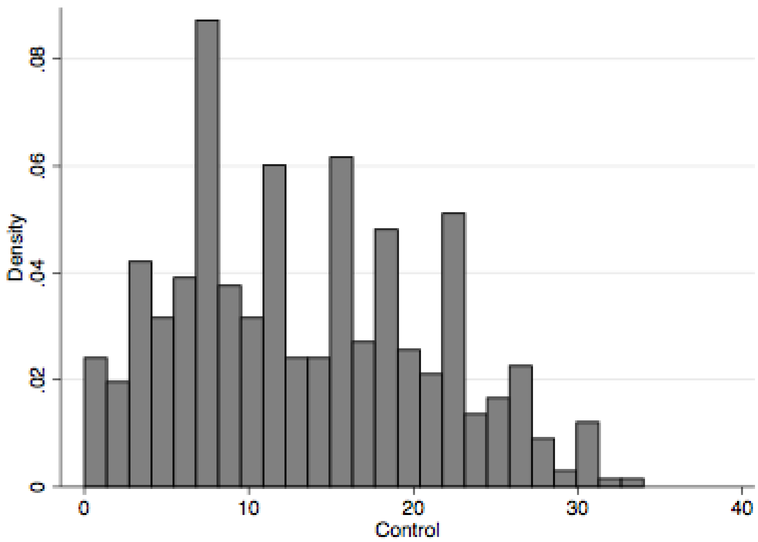
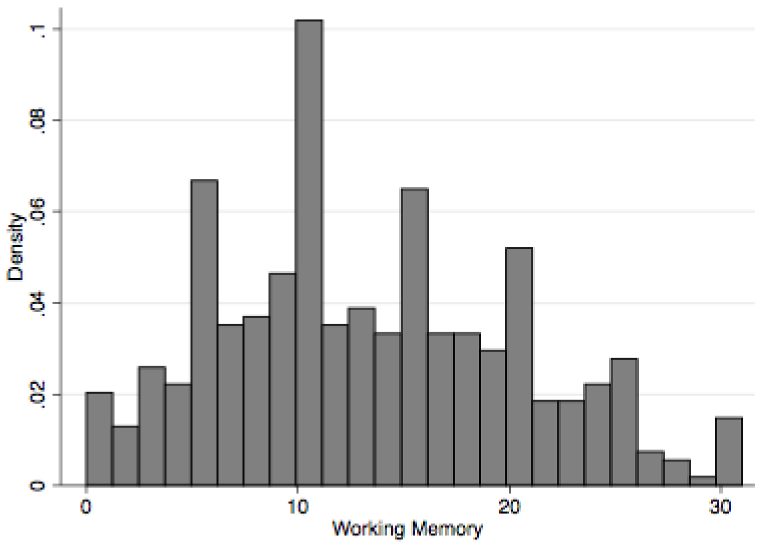


Figure 7: Histograms of KS1 Reasoning Scores by Group




Density
0
.02
.04
.06
.08
.1
0
10
20
30
Working Memory

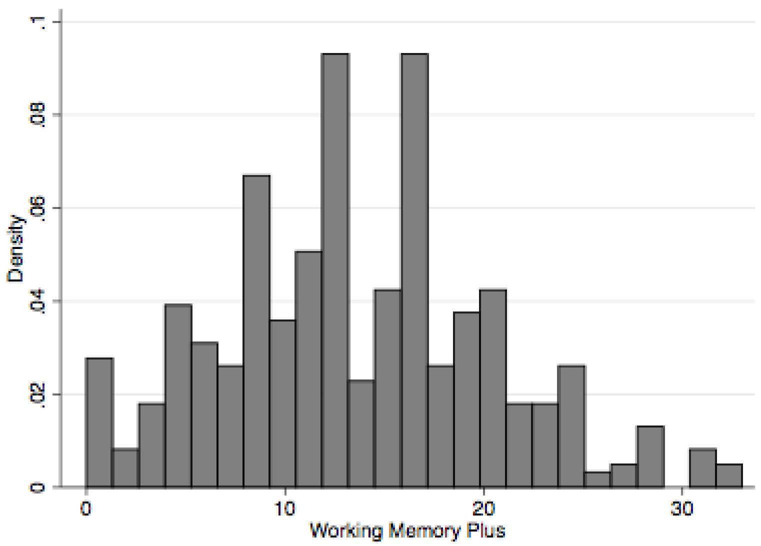
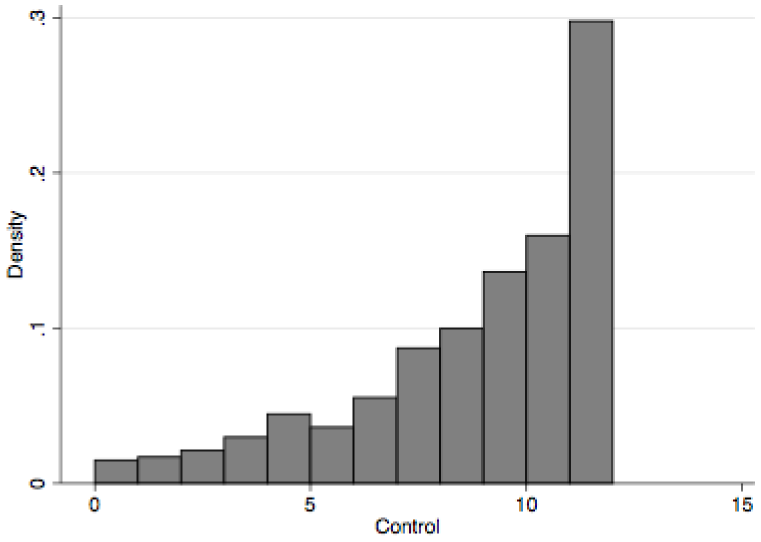
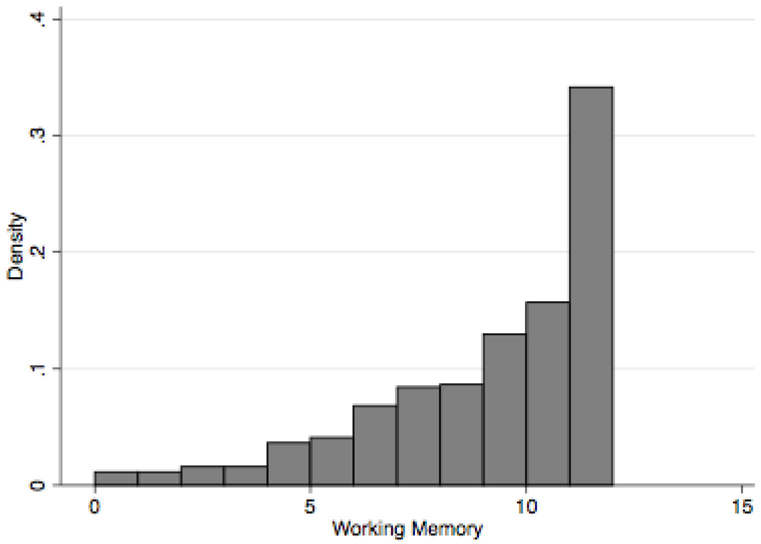
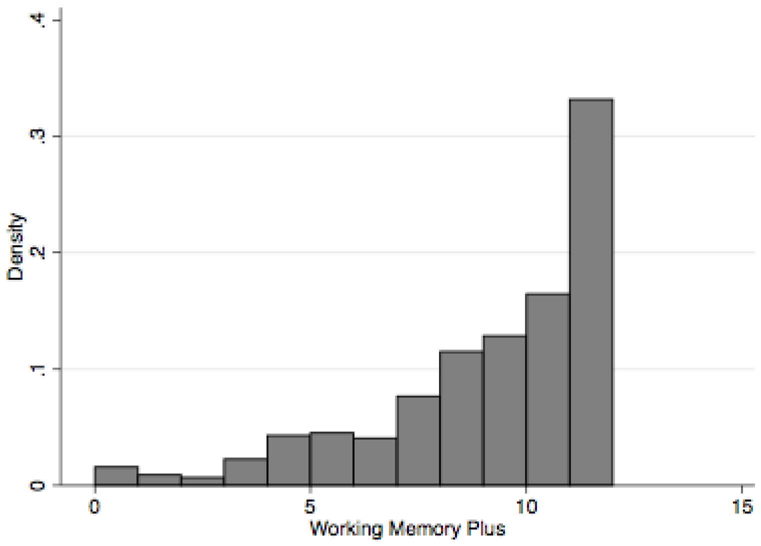

Density
0
.02
.04
.06
.08
.1
0
10
20
30
Working Memory Plus

Figure 8: Histograms of BAS3 Scores by Group






Density
0
.1
.2
.3
.4
0
5
10
15
Working Memory Plus

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

https://educationendowmentfoundation.org.uk

@EducEndowFoundn

Facebook.com/EducEndowFoundn