Read Write Inc. Phonics and Fresh Start

Evaluation Report

October 2022

Adria Molotsky, Paula Dias and Pooja Nakamura

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 Trust (now part of Impetus – Private Equity Foundation) 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

Contents

About the evaluator

The project was independently evaluated by a team from the American Institutes for Research:

Adria Molotsky, Paula Dias and Pooja Nakamura.

The co-principal investigators were Pooja Nakamura and Adria Molotsky.

Contact Details:

Dr Pooja Nakamura
American Institutes for Research
1000 Thomas Jefferson St NW
Washington, D.C. 20007

Tel: +1 (202) 403-5000
Email: pnakamura@air.org

Executive summary

The project

Read Write Inc. Phonics (RWI) and Fresh Start (FS) are a pair of programmes that aim to teach children to read and write through a systematic approach to literacy teaching. RWI is for children in Reception–Year 4 (age 4–9). FS is a catch-up programme for children in Years 5–8 (age 9–13) who have been identified as below their expected reading age. The programmes were developed by Ruth Miskin and delivered by Ruth Miskin Training (RMT). Queens University Belfast (QUB) recruited schools to this trial.

Both programmes start with the teaching of systematic synthetic phonics. Pupils work through a set of decodable books focused on letter–sound combinations pupils have already been taught and words that occur most commonly in written materials. The programmes progress to a focus on reading fluency and comprehension. RWI begins with 20-minute daily lessons in term 1 of Reception, building up to an hour a day for children in Year 1 and above. FS provides daily one-hour remedial lessons for 33 weeks, in place of or in addition to regular English lessons. Both interventions are delivered by teachers or teaching assistants. Schools in the intervention group for this trial received two regional training days for the headteacher and an appointed reading leader, followed by two whole-school training days, covering teaching techniques for both RWI and FS. In addition, schools received three to six development days per year in which trainers visited schools, and three leadership training days for the headteacher and reading leader during the first year of programme delivery.

131 schools took part in this effectiveness trial, including 4,914 pupils for the evaluation of RWI and 2,748 pupils for the evaluation of FS. The study was a two-arm blocked cluster randomised control trial (RCT). Randomisation was conducted at the school level, within blocks based on KS1 reading scores and Ofsted ratings. Schools in the intervention group were expected to deliver both RWI and FS. Pupils in Reception to Year 2 were expected to receive RWI while eligible pupils in Year 5 and Year 6 were expected to receive FS. The primary outcome for RWI was the New Group Reading Test (NGRT) for pupils in Year 2 at the end of the second year of the programme (who were in Year 1 in the first year of the trial). The primary outcome for FS was the KS2 reading fine points raw score, collected from two consecutive cohorts of pupils. The evaluation also included a survey administered to all intervention schools; interviews, focus groups and observations involving two treatment schools and one control school; observation of one teacher training day and interviews with three RMT staff members. The trial started in April 2016 and finished in July 2018.

Table 1: Key conclusions

Key conclusions
Children in Read Write Inc. Phonics (RWI) schools made the equivalent of one additional month's progress in reading, on average, compared to children in other schools. This result has a low to moderate security rating.
Children in Fresh Start (FS) schools made the equivalent of two months' less progress in reading, on average, compared to children in other schools. This result has a moderate security rating.
Children eligible for Free School Meals (FSM) in the RWI intervention group made the equivalent of three months additional progress in reading, on average, compared to children in other schools, while children eligible for FSM in the FS intervention group made the equivalent of three months less progress, on average. Both findings must be interpreted with caution, as they are based on small subgroups of pupils eligible for FSM.
RWI had a positive impact on pupils' phonics outcomes at the end of Year 1, equivalent to one month's additional progress. This was the average impact for pupils who received the programme for one year and pupils who received the programme for two years before taking the Phonics Screening Check. Neither children in the RWI or FS intervention group made more progress in writing than pupils in other schools.
FS was not implemented as intended in a significant proportion of intervention schools. 23 out of 66 intervention schools (35%) did not deliver FS at all and 19 schools (29%) delivered FS to some but not all eligible pupils. The low levels of engagement with FS will have influenced the impact results for this programme, so the results should be interpreted with caution. There was greater engagement with RWI, although 10 schools (15%) did not implement the programme.

EEF security rating

The findings for RWI have a low to moderate security rating and the findings for FS have a moderate security rating. This was an effectiveness trial, which tested whether the intervention worked under everyday conditions in a large number of schools. The trial was a well-designed two-arm RCT and was well-powered (a sufficiently large number of schools were recruited to the evaluation). However, there are factors that have reduced the security of the trial's findings. For RWI, 1,521 of the pupils who started the trial (31%) were not included in the final analysis due to pupil absence from school, moving to a different school, or teachers withdrawing pupils from testing due to concerns about the difficulty of the assessment. The evaluators also found that 13 schools (20%) in the RWI control group had either attended RWI regional trainings or purchased RWI materials, while 10 schools (15%) in the RWI intervention group did not deliver the programme. For FS, 466 of the pupils who started the trial (17%) were not included the final analysis, for reasons such as pupil absence from school or pupils moving to a different school. Additionally, 23 schools in the FS intervention group (35%) did not deliver the programme at all, while 19 schools (29%) delivered FS to some but not all eligible pupils and 8 schools (12%) did not provide enough data for the evaluators to know whether or not they were delivering FS. These factors make it harder to accurately estimate the size of the impacts of the programmes on pupils in the trial.

Additional findings

Pupils allocated to the RWI intervention group made, on average, one additional month's progress in reading compared to those in the control group equivalent, while pupils allocated to the FS intervention group made the equivalent of two months' less progress in reading, on average. These are our best estimates of impact. The findings for RWI have a low to moderate security rating while the findings for FS have a moderate security rating. As with any study, there is uncertainty around the results. The possible impact of RWI ranges from no additional progress to positive effects of up to two additional months progress, while the possible impact of FS ranges from negative effects of up to three months' less progress to negative effects of one month's less progress.

There is some evidence that, on average, RWI may have had a positive impact on pupils' reading but no impact on pupils' writing attainment. In the focus groups conducted in two schools for this evaluation, teachers and teaching assistants reported difficulties incorporating the writing component into RWI classes, which may explain this. The evaluation also found that RWI had a positive impact on pupils' phonics outcomes at the end of Year 1, equivalent to an average of one month's additional progress. Pupils who received the programme in both Reception and Year 1 had more positive phonics outcomes than those who received the programme only in Year 1, suggesting that pupils may benefit from receiving the programme across two academic years, as recommended by the programme provider.

Children allocated to the FS intervention group made the equivalent of two months' less progress in reading, on average, compared to the control group. However, there was some nuance within this result. Pupils in the cohort that were offered the programme for two years made no additional progress compared to the control group, while pupils in the cohort that were offered the programme for one year made 3 months' less progress in reading. FS was not implemented in 35% of intervention schools and only implemented to some eligible pupils in 29% of intervention schools, which will have influenced these results. Barriers to the implementation of FS identified in interviews and focus groups with staff from two schools included limited space or staff capacity for delivering the programme, the lack of FS-specific training provided to teachers and the FS handbook being less comprehensive than the RWI handbook. Programme staff from RMT interviewed for the evaluation also reported that a lack of clarity on the specifics of the FS programme during the initial recruitment of schools by a third party led to implementation issues later on. Pupils in both the intervention and control group for FS were identified as below their expected reading age, and data collected at the start of the trial suggests that a variety of programmes and approaches were being used in control schools to support low-attaining readers, as part of usual practice in these schools.

There is extensive evidence in the wider literature on literacy that systematic synthetic phonics approaches can have substantial positive impacts on pupils' reading outcomes. The results of this evaluation suggest that being in the RWI intervention group may have contributed to gains in pupils' reading, while being in the FS intervention group may have contributed to negative impacts on pupils' progress in reading. This contrasts with the results of the EEF's previous efficacy trial of FS, which identified a positive impact equivalent to three months' additional progress in pupils' reading, on average. Differences between this trial and the previous FS trial include the implementation challenges and limited school engagement with the programme observed in this evaluation, which were not observed in the previous trial, as well as differences in the evaluation design. The 10 schools in the previous trial ran their own evaluations of FS with

support from a team of evaluators, who aggregated results across schools. The previous trial also used a different primary outcome measure for FS (the NGRT instead of KS2 reading scores) and delivered to a different year group (Year 7 instead of Year 5).

Cost

The average costs of RWI for one school was around £18,960, or £186 per pupil per year when averaged over 3 years. Costs varied based on school Ofsted ratings, because schools with low Ofsted ratings received additional in-school development days compared to schools with high Ofsted ratings.

Impact[1]

Table 2: Summary of impact on primary outcome

Outcome/Group	Effect size (95% confidence interval)	Estimated months' progress	EEF security rating	No. of pupils	p-value	EEF cost rating
RWI: NGRT SAS reading score – Year 2	0.05 (−0.02, 0.12)	1	🔒🔒🔓🔓🔓	3,393	0.528	£ £ £ £ £
RWI NGRT SAS reading score – Year 2 FSM subgroup	0.22 (0.08, 0.37)	3	N/A	797	0.058	N/A
FS: KS2 reading score – cohort 1 (Year 7 at end of trial)	−0.22 (−0.33, −0.11)	−3	🔒🔒🔒🔓🔓	1198	0.015	N/A
FS: KS2 reading score – cohort 1 FSM subgroup (Year 7 at end of trial)	−0.26 (−0.43, −0.10)	−3	N/A	599	0.014	N/A
FS: KS2 reading score – cohort 2 (Year 6 at end of trial	−0.03 (−0.15, 0.09)	0	🔒🔒🔒🔓🔓	1120	0.699	N/A
FS: KS2 reading score – cohort 2 FSM subgroup (Year 6 at end of trial	−0.13 (−0.30, 0.03)	−2	N/A	544	0.157	N/A
FS: KS2 reading score – combined	−0.13 (−0.21, −0.05)	−2	🔒🔒🔒🔓🔓	2318	0.090	N/A
FS: KS2 reading score – combined FSM subgroup	−0.21 (−0.32, −0.09)	−3	N/A	1143	0.016	N/A

[1] This work was produced using statistical data accessed via the ONS Secure Research Service. The use of this data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates.

Introduction

Background

According to the National Curriculum Assessment (2019), about 30% of pupils leave primary schools without achieving the expected attainment level in reading (DfE, 2019a). In 2019, 82% of pupils met the expected phonics screening check in Year 1, while by the end of Year 2, approximately 91% of pupils met the standards, a 1 percentage point decrease from 2018. Key Stage 1 (KS1) reading scores remained at 75%, which reflected no change from 2018 (DfE, 2019b). Attainment in reading has been an important focus for the Department for Education in England, and there have been various policies implemented in order to achieve the national targets of attainment in reading at Key Stage 2 (KS2). The government introduced pupil-premium funding for schools to raise the attainment level of disadvantaged pupils. Schools now have the opportunity to use these funds to purchase resources or teaching approaches that can support pupils who are at risk of falling behind. An essential step is to identify reading approaches that ensure all pupils meet National Curriculum expectations, including those pupils who are at risk of falling behind.

Large-scale reviews of the effectiveness of reading interventions in the United Kingdom, the United States and Australia have shown that phonics instruction is significantly more effective than non-phonics approaches for teaching reading (AGDEST, 2005; Chambers et al., 2015; NICHHD, 2000; Torgerson et al., 2006). In the United Kingdom, the Rose Report underscored specifically that the most effective kind of systematic phonics approach is synthetic phonics (Rose, 2006).

According to many theories of reading development, both good word reading skills (defined as the ability to automatically recognise printed words and sound them out accurately) and language comprehension skills (defined as the ability to understand and use the oral language, including syntax, semantics and other linguistic subskills) are considered required for reading comprehension development; neither is sufficient on their own (Gough & Tunmer, 1986; Hoover & Gough, 1990; Hoover & Tunmer, 2020; Kim, 2017).

While there is some evidence for the effectiveness of phonics approaches for older pupils with reading difficulties (Ehri et al., 2001), most of the studies examine the effectiveness of phonics on younger pupils, especially at the Early Years Foundation Stage and KS1. Furthermore, there is less evidence on spelling and reading comprehension outcomes (as opposed to word reading). More evidence is necessary to understand what aspects of reading development are supported by a phonics-focused approach for older pupils.

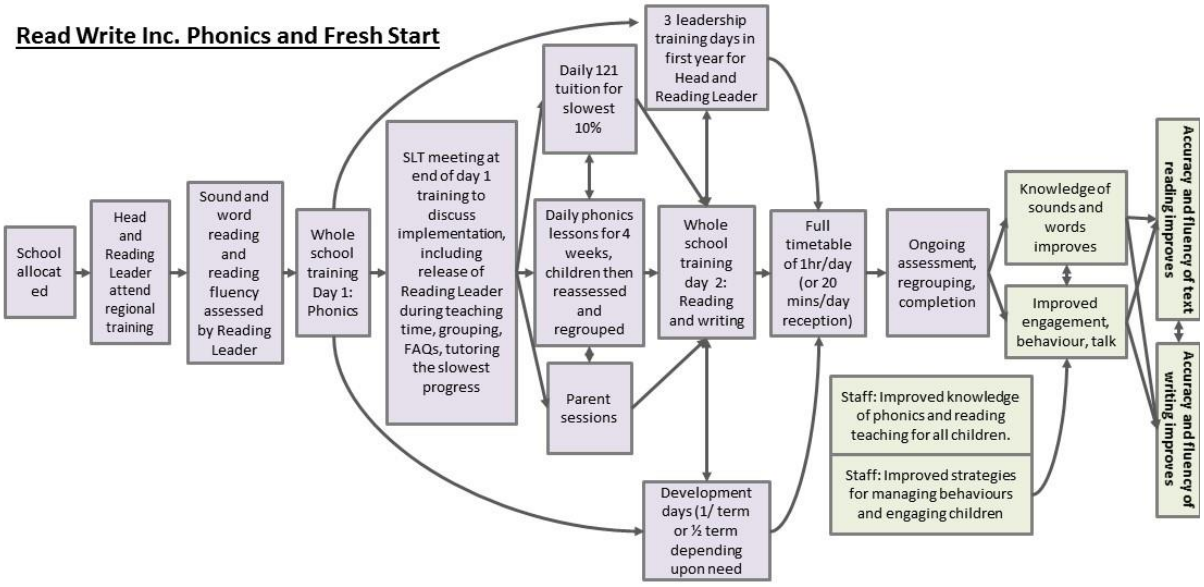
Previously, Fresh Start (FS) was evaluated in an aggregated efficacy trial, which showed promising results as remedial reading intervention for Year 7 pupils (Gorard et al., 2015). Given that this evaluation was conducted independently by schools, there was the potential for selection bias (although the independent evaluators coordinating the efficacy trial did not see any evidence of such bias), and there was also inconclusive evidence about the programme's effectiveness under diverse conditions. When we designed this evaluation, we aimed to address these concerns by conducting our own randomisation and examining the degree to which the programme works with differing levels of support from trainers, heterogeneity in the groups of pupils, and the duration of the programme.

Intervention

Read Write Inc. Phonics (RWI) and FS are systematic synthetic phonics-based literacy programmes. RWI is a complete literacy programme targeted at 4- to 9-year-olds (Reception to Year 4) learning to read and write as well as those who need to catch up. RWI is delivered to pupils by teachers or teaching assistants in groups based on reading level; at the beginning of the year, pupils are assessed and divided into groups by ability level to ensure they are being taught at the appropriate level. Children making the slowest progress with RWI additionally receive daily one-to-one tutoring. The RWI programme starts off with 20 minutes of daily lessons in term 1 of Reception and builds to 40 minutes per day by the end of the year. Meanwhile, pupils in Year 1 and above receive 1 hour of RWI lessons per day. FS is a 33-week catch-up literacy programme delivered by teachers or teaching assistants in a small group setting to target children in Years 5–8 (9 to 13 year-olds) identified as having fallen behind in expected standards in reading and writing; it aims to support pupils with reading difficulties at the end of primary school and beginning of secondary school, using systematic phonics approaches embedded in content that is tailored for older pupils. The programme recommends a FS session every day instead of, or in addition to, the usual English lessons, and each FS session should be one hour long. The present evaluation focused on pupils in Reception to Year 2 for RWI and pupils in Years 5 and 6 for FS.

As part of this project, both programmes involved the same training (depicted in Figure 1), which was usually delivered to the whole school by Ruth Miskin Training facilitators. The headteacher and reading leader also attended a regional training prior to the whole school training. After this, the reading leader assessed the children on sound and word reading and fluency using a short, five-minute assessment. The first day of the whole school training was for all teaching staff including the headteacher and focused on the phonics element of the programmes. Three or four weeks later, the trainer returned to support all staff on a normal teaching day. She provided feedback, further training and coaching with a focus on the slowest progressing children. The second day of the whole school training was approximately four to six weeks after the first training day and focused on teaching reading and writing. In the first year of implementation the reading leader and headteacher attended two additional leadership training days, for a total of three leadership training days. In the second year of implementation, they attended an additional leadership meeting. The trainer continued to visit the school termly or half-termly (more visits were conducted for schools with lower Ofsted ratings) to support and coach staff and work closely with the reading leader.

Figure 1: RWI and FS training and support



The FS programme showed positive results through an EEFs aggregated trial, which involved 10 schools running their own within-school RCTs. The study we conducted is an RCT where randomisation was conducted at the level of the school. Thus, while the aggregated trial used a non-standard RCT methodology, this current study serves as an effectiveness trial with implementation conducted at-scale. However, both the previous study as well as the current study of FS have both strengths and limitations which will be further discussed in the *Limitations and lessons learned* section of the *Conclusion*. Furthermore, in addition to exploring results for FS, this trial also explored results for RWI. With the RCT design, we determined the impact of the programmes on various pertinent reading outcomes.

Evaluation objectives

The main research question for the evaluation was as follows:[2]

1. Do RWI and FS impact pupils' reading outcomes?

This research question focused on pupil reading while the impact of the programme on writing was also investigated. Both RWI and FS are intended to help teachers boost pupil reading and writing skills. Hence, the ultimate outcome would be a meaningful increase in pupils' reading and writing skills.

Although the evaluation's primary outcomes concerned the programmes' impact on pupils, both the RWI and FS work by changing the way in which teachers approach phonics and reading lessons. As such, the process evaluation focused on the following questions related to teachers and implementation:

2. How are the programmes being implemented?
3. Are the teachers using the programmes as envisioned by the developers?
4. Are the teachers receiving support from the reading leader?
5. How does training quality help support the quality of the reading leader to implement the programme with fidelity?
6. How contingent is the success of the programme on this continuous professional development?
7. How has the programme impacted teacher knowledge on how to teach reading?
8. Does the programme change teacher practices and behaviours in the classroom?

Ethics and trial registration

- QUB obtained approval through their ethics board for all the data collection and analysis and applied to the National Pupil Database (NPD) for access to pupil outcome and background data. Since NPD data cannot leave the UK, evaluators from AIR became visiting fellows at QUB and were covered under QUB's Ethics Board approval. The American Institutes for Research (AIR) accordingly obtained exempt status from its Institutional Review Board (IRB), which is registered as a research institution (IORG0000260) with the US Office of Human Research Protection at the US Department of Health and Human Services.
- The activities undertaken by QUB and AIR were also scrutinised by the QUB School of Education Ethics Committee by three independent academic reviewers. The project was approved by the QUB School of Education Ethics Committee prior to any data being collected. Any participants (in this case schools acting as in parental locus, and parents) were provided with clear information on data collection, processing and retention. Individuals were informed that they had the right to withdraw their consent to us processing data, and we would comply with this to comply with our ethical processes, up until the point that their data is anonymised and aggregated. Schools signed Memorandums of Understanding (MoUs) that set out the roles and responsibilities of QUB and the schools. Participants were informed that data would be stored for a minimum period of five years by QUB. Participants provided opt-out consent, but only one pupil opted out. Examples of the information on data collection (provided in June 2016, March 2018 and again in July–September 2018), MoUs (signed by schools in Feb–June 2016) and consent forms (provided in June 2016 and again in July–September 2018) are contained in Appendix F.
- The study has been registered and the International Standard Randomised Controlled Trial Number (ISRCTN) is ISRCTN97725862.

Data protection

AIR, as one of the study's data controllers, takes the protection of data seriously and has implemented numerous companywide policies to promote the security of data. Staff are required to participate in yearly training on data security.

[2] The statistical analysis plan (SAP) can be found on the effectiveness trial website: https://educationendowmentfoundation.org.uk/public/files/Projects/Fresh_Start_SAP_2018.01.17.pdf

All AIR computers are encrypted, and password protected with stringent requirements for passwords, including specification on password length, character usage and password expiration dates. Any login into AIR's servers and systems (including accessing email or shared network) outside of AIR offices require duo authentication.

- In accordance with the Data Protection Act and GDPR, for this evaluation, no data were stored on AIR's computers. The lead evaluator became a visiting fellow of QUB, the other data controller and the data processor for this evaluation and analysed the data while in the UK from the Secure Research Service through the Northern Ireland Statistics and Research Association. Additionally, the lead evaluator from AIR became a UK Office of National Statistics Accredited Researcher along with all researchers on the team from QUB.
- The data used in this evaluation included information on age, gender, ethnicity, reading attainment scores and KS1/KS2 assessments, all of which was obtained from NPD. Data obtained through the NGRT assessment was processed by QUB researchers, who assisted with the administration of the assessments in primary classrooms. However, names of pupils were not collected or stored at any stage and only unique pupil numbers (UPNs) were used to match data in different phases of the project. We consulted with data processors at QUB, to help us identify risks of the data processing activities associated with the project. As with most university-based research, we had two parallel processes. One involved the processing of data ethically, and one involved the processing of data in compliance with appropriate legislation. This project started under the Data Protection Act legislation, but in May 2018 new legislation, GDPR, came into effect, meaning that a different framework was required for the final phase of data processing.
- The legal basis for processing and holding data on the basis of public task basis has a sound basis in law. Under QUB's Charter and Statutes its object is the "advancement and dissemination of learning and knowledge through teaching and research…". As such the processing of personal data for research within the University is an underlying function with a clear basis in law. In addition, Article 6(1)e of the GDPR was used as the lawful basis for processing personal data as part of this project. This is generally known as the "public task" basis. Further, the basis of "research" as the lawful means by which special data is processed including ethnicity and ever FSM status (Article 9(2)j of the GDPR). Current ICO guidance available here: https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/public-task/, was reviewed and determined, in collaboration with the Data Compliance Office at QUB, that this research formed part of its performance of a task in the public interest, as one of the core purposes of QUB provided for in its Charter and Statutes. Although data outlined above had been collected prior to GDPR coming into effect, further communication was sent to parents in July–Sept 2018 to outline the data that was being collected and stored (see forms F.4, F.5, F.8 and F.9 in Appendix F for the updated information and consent forms for RWI and FS, respectively).
- Data was screened to assess the risk of unintentional disclosure. Thurston, Cockerill, Mulgrew and O'Neil (QUB) all completed safe researcher training and assessment provided by NISRA and Thurston completed the ESRC version of this safe data handling course/assessment. The team have been working in England for a number of years. They had already developed safe procedure for transferring data through encrypted point to point networks (QUB developed a University housed/maintained dropbox to facilitate this). No data was stored nor passed without encryption with non-dictionary complex passwords. All measures such as these were incorporated into any work plan that this team used during the study. At no point was data processed that included names of pupils, only UPNs being used to match data. The risk of unintentional disclosure was therefore assessed as low.

Project team

Ruth Miskin Training (RMT) team

The Ruth Miskin Training team included Ruth Miskin, who developed the RWI and FS programmes, as well as individuals who trained the teachers, and managed the overall implementation of delivery from developers to trainers to teachers. The Ruth Miskin Training team was supported by Oxford University Press, the publisher of RWI and FS programmes. Oxford University Press also provided a discount to schools partaking in the trial with respect to the purchasing of the RWI and FS programmes.

Ruth Miskin: Founder and CEO

Mel Erwin: Director

Nina Coglin: Head of Operations

Evaluation team: American Institutes for Research (AIR)

AIR was the lead evaluator responsible for the overall trial design, all data analysis and authoring this evaluation report.

Pooja Nakamura, PhD: Principal Investigator

Adria Molotsky, PhD: Principal Investigator and lead evaluator

Paula Dias, PhD: Head of qualitative analysis

Recruitment and data collection partner: Queens University – Belfast (QUB)

QUB was responsible for collecting data and agreements from schools at the time of recruitment, administering the pupil NGRT assessment and requesting data from the NPD.

Allen Thurston, PhD: Project Director

Maria Cockerill, PhD: Senior Research Fellow and head of logistics

Nicole Craig, PhD: Researcher

Nina O'Neill, PhD: Researcher

Method

Trial design

The study was a two-armed cluster RCT with treatment assignment at the school level. Because the RWI intervention was delivered to all 4- to 6-year-olds, randomisation was conducted at the school level rather than the classroom level to prevent spillover (contamination) of the intervention from treatment to control classrooms that might be in the same school and enable treatment-group teachers to collaborate with their colleagues after receiving training on RWI. Schools allocated to the intervention group for RWI were also expected to deliver FS to eligible Year 5 and 6 pupils. These pupils formed the intervention group for FS. We conducted stratified randomisation using strata based on the school-level historical KS1 reading score (provided by schools as part of the trial registration form) and government-assigned Ofsted grades to ensure an even spread of schools by performance to each treatment group. For the historical KS1 reading score, we grouped schools according to quartiles, while we assigned the schools two groupings within the Ofsted grade – those with a high grade (1 or 2) and those with a low grade (3 or 4).

To facilitate Queen's University Belfast's (QUB's) recruitment of schools, randomisation was conducted in stages, with the first batch of schools randomised in April 2016, the second batch in May 2016, the third batch in June 2016 and the final batch in July 2016. For the second, third and fourth batches of schools, we assigned quartile levels of KS1 reading scores based on the values of the quartiles for the first batch of schools.

Table 3: Trial design

Trial design, including number of arms		Two-armed, cluster randomised controlled trial (RCT)
Unit of randomisation		Schools
Stratification variable (s) **(if applicable)**		KS1 reading scores grouped into quartiles Government assigned Ofsted grades: high (grade of 1 or 2) and low (grade of 3 or 4)
Primary outcome	**Variable**	RWI: New Group Reading Test (NGRT) is a standardised assessment that reliably measures reading skills such as sentence completion and comprehension skills. FS: National KS2 reading assessment (commonly referred to as SATs) measures pupils' reading comprehension in relation to the National Curriculum standards.
	Measure (instrument, scale, source)	RWI: NGRT standardised age score ranging from 70 to 130 points, Grade Level 3 FS: KS2 reading score ranging from 0 to 50 points, NPD
Secondary outcome(s)	**Variable(s)**	RWI: National KS1 writing test assessed by teachers measures pupils' ability to meet the objectives for a piece of writing in relation to national standards. Year 1 phonics screening check is a way for teachers to ensure pupils have sufficient phonic knowledge to decode words accurately. FS: National KS2 writing test administered by teachers measures pupils' ability to meet the objectives for a piece of writing in relation to national standards.
	Measure(s)	RWI: KS1 writing percentage pass rate, NPD; Year 1 phonics screening check scores ranging from 0 to 40 points, NPD

[3] The evaluation team originally planned on using the raw scores from the NGRT as the primary outcome for the RWI analysis. However, upon receipt of the NGRT, we realised raw scores were not reported, rather standardised age scores were reported. Thus, we use the standardised aged scores as our default primary outcome.

	(instrument, scale, source)	FS: KS2 writing scores ranging from 0 to 70 points, NPD.
Baseline for primary outcome	**Variable**	RWI: none available FS: KS1 reading scores
	Measure **(instrument, scale, source)**	FS: Teacher-delivered national reading assessment with scores ranging from 0 to 40 points

The trial was designed as a cluster RCT to determine the effectiveness of the programmes on various pertinent reading outcomes (see Table 3 for specifics about the outcome measures). The RCT has two arms and compared the reading achievements of pupils in schools who received the programmes to pupils in schools who form a two-year wait-list control. The control schools continued 'business-as-usual' for the teaching of phonics, reading and writing. That is, the treatment arm included those schools that received RWI and FS programmes, and the control arm included those schools who would not initially receive either programme but would become eligible to receive RWI and FS programming after the trial completed.

Participant selection

The QUB team recruited schools through their networks from different areas of the country from April to July 2016. The aim of recruitment was to generate a sample of schools across a wide range of performance while over-recruiting underperforming, disadvantaged schools. To be eligible, schools had to be state funded, and a large (but no more than 50%) share of the schools were targeted to come from the North East region of the UK as this was part of the EEF's North East Primary Literacy Campaign. In our final sample, 22% of schools came from the North East region of the UK.

To recruit schools, QUB sent an initial communication including a flyer from RMT and information about the trial to school headteachers. Once schools expressed interest, the schools completed a registration form. The form included information on school characteristics such as percentage of FSM, geographic region, school size, class size, gender ratios and Ofsted category. The information provided on the registration form was used in the random assignment of the schools, as described. In the end, QUB successfully recruited 135 schools, 68 of which were initially randomised to receive treatment.

While RWI was delivered to all children in Reception and Year 1 and children in Year 2 to Year 4 assessed as needing it, the evaluation focused on pupils from Reception–Year 2 only. While FS targeted eligible pupils in Years 5–8 who are below appropriate reading age, the evaluation focused on eligible pupils from Years 5 and 6 only (only pupils in primary school). Table 4, below, shows the grade level of pupils at the randomisation stage and then at the end of the intervention. Throughout this report, we refer to pupils by the grade they were in at the end of the evaluation in summer 2018: Years 1–3 for RWI and Years 6–7 for FS. It should be noted that, in some cases, the year group pupils were in at the end of the evaluation is not the year in which they received the intervention or completed the outcome testing. For example, the cohort referred to in this report as the 'Year 7 FS cohort' received the intervention and completed outcome testing during the first year of the intervention delivery, at which point they were in Year 6. At the end of the evaluation, they were in Year 7, but they did not receive FS or take part in outcome data collection in Year 7. Please see Table 4 and Figure 2 for more information.

The eligibility criterion for the FS programme was a school progress measure which was consistent across all schools, through which the children were selected according to their reading levels. Children reading below age-related expectations were expected to be taught FS. The programme was targeted to those children who were not on track to meet national expectations in reading at the end of KS2. These were pupils whose reading age was 9 years or below, which is 1 to 2 years below expectations. While the school knew the reading age compared to the actual age of the pupils, each pupil's progress towards 4b achievement was assessed by the teacher.

Table 4: Grade level at randomisation vs end of intervention

Grade level at randomisation (summer 2016)	Grade level at end of intervention (summer 2018)
RWI	
Nursery	Year 1
Reception	Year 2
Year 1	Year 3
FS	
Year 4	Year 6
Year 5	Year 7

Outcome measures

For both programmes we expected to see a positive increase in the primary and secondary outcomes measured, which are described in more detail below. We used the SAS for NGRT and raw scores for the other primary and secondary outcomes, where the raw score was the total number of marks a pupil received on a test, according to how many questions they answer correctly, or the writing score variable from NPD.

Primary outcomes

RWI primary outcome: The primary outcome was the NGRT standard age score (SAS)[4] results at the end of the second year of the programme (the NGRT was only administered to pupils in Year 2 during the 2017/18 school year) that was electronically administered by QUB. The NGRT is a standardised, adaptive assessment that reliably measures pupil reading skills. The classroom teachers were trained on the administration of the tablet-based version of the NGRT by QUB staff prior to pupil assessment. No issues were reported to AIR with respect to the administration of the survey. However, some teachers voiced concerns over the perceived difficulty of the assessment.

FS primary outcome: The primary outcome was KS2 reading fine points raw score from the NPD for eligible pupils who sat this assessment at the end of the first year of programme implementation (these pupils took the assessment in Year 6 but are referred to in this report as the Year 7 cohort, based on their year group at the end of the evaluation) and eligible pupils who sat the assessment at the end of the second year of programme implementation (these pupils are referred to as the Year 6 cohort). We examined this outcome for each cohort separately as well as combined for improved precision, which is the primary outcome.

Secondary outcomes

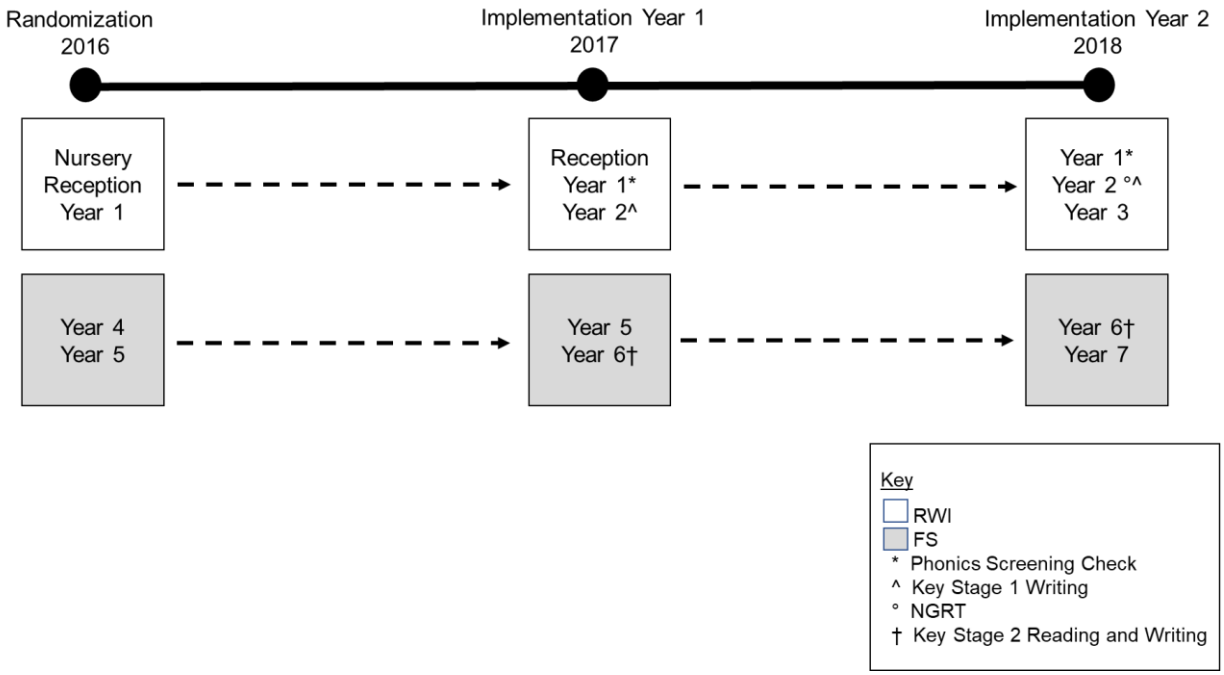
RWI secondary outcomes: The secondary outcomes included the teacher-assessed KS1 writing percentage pass rate at Year 2 and the teacher-assessed Year 1 phonics screening check for both years of the programme from the NPD. We analysed the data separately for each year of the programme as well as aggregated across years.[5] Figure 2 shows when each of the year groups were assessed for each secondary outcome. We examined the percentage of children that passed the KS1 writing assessment and both the percentage of children that passed the phonics screening check as well as the raw score of the phonics screening check. Because the teachers had knowledge of the treatment status when they conducted the Year 2 KS1 Writing and Year 1 Phonics screening check, these outcomes are likely to be more biased than the NGRT measure. The Department of Education provided standardised guidelines for the implementation of these tests and thus helped minimise bias.

[4] The SAS uses the pupil's raw NGRT score and standardises using the scores of other pupils of their age level. The score is then placed on a scale, with an average of 100, which allows for a comparison with a nationally representative sample of pupils of the same age across the UK. We use the SAS for this analysis rather than raw scores as NGRT does not report raw scores.
[5] The year group refers to the year the pupils were in at the end of the evaluation for each primary and secondary outcome.

FS secondary outcome: The secondary outcome was the combined KS2 writing results for both years of the programme from the NPD. This was the combined results for the Year 7 cohort (who received FS and completed outcome testing while in Year 6 in 2017 but were in Year 7 when the evaluation ended in 2018) and the Year 6 cohort (who began receiving FS while in Year 5 and completed outcome testing at the end of Year 6 in 2018).

Figure 2 shows the timeline by cohort of when each examination was administered to pupils for ease of comprehension.

Figure 2: Outcomes by year and grade



RWI outcome data for this evaluation included: the Phonics Screening Check results from pupils in Year 1 in summer 2017 and pupils in Year 1 in summer 2018; KS1 writing results from pupils in Year 2 in summer 2017 and pupils in Year 2 in summer 2018; and NGRT results from pupils in Year 2 in summer 2018. For FS, outcome data in this evaluation included: KS2 reading and writing scores for pupils in Year 6 in summer 2017 and pupils in Year 6 in summer 2018.

Sample size

For this study, we used a multilevel model with two levels (pupils and school) with treatment occurring at the second level (the school level). We selected this model based on schools being the unit of randomisation, with pupils being the unit of observation.

To establish the necessary sample size for our study, we conducted statistical power calculations using the PowerUp! tool (Dong and Maynard, 2013). The following parameters were used in our calculations:

- Alpha level (α) = 0.05, two-tailed test;
- Power (1 – β) = 0.80;
- Intraclass correlation (ICC) = 0.10 for RWI and 0.13 for FS;
- Proportion of Level 2 units randomised to treatment = 0.50;
- Number of pupils per school: 37 for RWI, of which 11 would be FSM pupils and 10 for FS, of which 3 would be FSM pupils;
- Proportion of pupil variance explained by pupil covariates (R^2): 0.00 for RWI and 0.53 for FS.

The selection of the alpha and power parameters are standard values used in RCTs. The ICCs for each intervention were from the 2013–2014 NPD[6] while the proportion of pupil variance explained by pupil covariates for FS was set in discussions with the project team.[7] As we did not access pupil covariates for RWI, the proportion of variance explained by these covariates is inherently 0.00.

The initial sample size for the evaluation was calculated based on a 120-school RCT, with an even split between the treatment group and the control group. We determined that a sample of 120 schools split into treatment and control arms in equal proportion would be sufficient for the study. For RWI, the sample would consist of 4,400 Year 2 pupils (37 pupils per school), and for FS the sample would consist of 1,200 Year 6–7 pupils (10 pupils per school with 5 from each year group).[8] The actual number of schools used for the intention-to-treat (ITT) analysis is 131, higher than anticipated. While the average number of pupils per school remained the same for RWI at the recruitment stage, the average number of FS-eligible pupils per school was larger; on average, the schools had 21 pupils eligible for FS, of which 10 were eligible for FSM.

In the impact evaluation section, below, we present the calculations of the minimum detectable effect size (MDES) and comparisons across the initial sample size, those based on the actual number of schools in the study and those based on the analytic sample (see Table 6).

Randomisation

We conducted stratified randomisation at the level of the school in stages to ensure an equal distribution of schools by performance level prior to the implementation of the intervention. The stratification created groups based on the school-level historical KS1 reading score provided by schools as part of the registration form that QUB collected and the government assigned Ofsted grade, a measure of school quality that determines what kind of training they receive. For the historical KS1 reading score, we grouped schools according to quartiles. We assigned the schools into two groupings within the Ofsted grade, those with a high grade of 1 or 2 and those with a grade of 3 or 4.

AIR's team of researchers conducted the randomisation using Stata to randomly assign numbers to schools within each stratum. Schools were then ordered from largest to smallest, based on the randomly generated number and the top 50% within strata were selected as intervention schools.

To facilitate QUB's collection of UPNs, randomisation was conducted in stages, with the first batch of schools in April 2016, the second batch of schools in May 2016, the third batch of schools in June 2016, and the last batch of schools in July 2016. For the second, third and fourth batches of schools, we assigned quartile levels to KS1 based on the values of the quartiles for the first batch of schools. After randomisation was completed, the study team confirmed the quartiles from the first batch were valid for the following batches as well.

[6] Intracluster correlation was informed by the ICCs for KS1 and KS2 for RWI and FS, respectively, from the 2013–2014 NPD available at https://educationendowmentfoundation.org/pulic/files/Evaluation/Writing_a_Protocol/ICC_2015.pdf.
[7] The proportion of variance in outcomes explained by pupil level covariates was agreed upon during the project's second set up meeting on 5th January 2016.
[8] The year group refers to the year pupils were in at the end of the evaluation, in summer 2018. The RWI pupils referred to as the Year 1 cohort were in Year 1 when the evaluation ended but began receiving the intervention while in Reception and completed outcome testing in Year 1. The RWI pupils referred to as the Year 2 cohort were in Year 2 when the evaluation ended but began receiving the intervention while in Year 1 and completed outcome testing in both Year 1 and Year 2. The RWI pupils referred to as the Year 3 cohort were in Year 3 when the evaluation ended but received the intervention and completed outcome testing while in Year 2. The FS pupils referred to as the 'Year 7 cohort' were in Year 7 when the evaluation ended but received the intervention and completed the outcome testing while in Year 6. The FS pupils referred to as the 'Year 6 cohort' were in Year 6 when the intervention ended. They began receiving the intervention while in Year 5 and completed outcome testing in Year 6.

Statistical analysis

Primary analysis

We analysed the ITT impact of the programmes using the following difference-in-means specification. All analyses were ordinary least squares regressions conducted at the individual pupil level using the statistical software package Stata accounting for the nested structure of the data with clustered standard errors.

$$Y_{ijk} = \alpha + \beta R_k + \theta X_{ijk} + \varepsilon_{ijk} \quad (1)$$

where Y_{ijk} is the outcome measure for pupil i in classroom j in school k. RWI was administered to all children in Reception and Year 1, so the pupils in the primary ITT analysis were all pupils in Reception at the time of randomisation from the study schools (pupils who were in Year 2 at the end of the intervention and able to take the NGRT). On the other hand, FS was targeted at eligible pupils, so the pupils in the ITT analysis were all pupils identified as eligible by both intervention and control schools prior to randomisation. That is, when analysing the programmes' effects for pupil i in classroom j in school k, Y_{ijk} represents the primary outcomes described in more detail above, specifically for FS the combined[9] Year 6 and Year 7 KS2 reading results and for RWI the NGRT SAS at the end of the second year for Year 2 pupils. R_k is a variable that equals one if the pupil's school was assigned to the treatment (which includes both FS and RWI) and equals zero if not. X_{ijk} is a vector of controls.

For our primary specification, the vector of controls included the baseline measure of the outcome variable so that we can perform ANCOVA analysis, which improved our power.[10] The vector of controls also included the groups created for the stratified randomisation based on the school-level historical KS1 reading score and the Ofsted grades. Due to the randomised treatment assignment, including additional covariates should have no effect on the point estimates of the programmes but may be useful for statistical precision. Thus, we conducted a secondary specification that included the following variables from the NPD: gender, ethnicity (as defined by the schools), whether the pupil speaks English as an additional language (EAL), and an indicator for whether the pupil was ever eligible for Free School Meals (FSM) in the past 6 years, which corresponds to EVERFSM_6_p_ in the NPD.[11]

In addition to these two specifications, we estimated results for the standard EEF model (EEF, 2018). In this EEF standard model, the vector of controls included only the baseline measure of the outcome variable. Thus, the covariates for this secondary specification are the treatment group status and baseline measure of the outcome variable only. However, for RWI no baseline measure of the outcome variable exists. For RWI the standard EEF model included only the covariate of the treatment assignment.[12]

With the ordinary least squares regressions at the individual pupil level, we used Huber–White cluster-robust standard errors in our estimates. The use of cluster-robust standard errors accounts for a lack of independence across observations due to the clustering of pupils at the level of school. Doing so ensures that standard errors are properly estimated. β is our coefficient of interest, representing the marginal effect of being part of the RWI or FS programmes, as compared to the control group.

[9] As discussed in the primary outcome section for FS, we will examine this outcome for each cohort separately as well as combined for improved precision. However, the primary outcome will be the combined variable for improved statistical precision.
[10] ANCOVA analysis is not possible with the analysis for RWI since no baseline measure of the outcome variable exists. Since the pre-test is not available for RWI, the secondary specification with additional controls could be very useful for efficiency. In the analysis for FS, the covariates will include the KS1 reading results as a baseline measure of the outcome variable. It is worth noting that the KS1 tests are teacher administered so they could contain some bias. However, since the KS1 tests will have been conducted prior to randomisation, the risk of bias is decreased. Nevertheless, for the teacher-assessed KS1 results, we will first normalise the scores to bring all scores into proportion with each other.
[11] We did not conduct these analyses for pupils who were in Nursery at the time of randomisation as we do not have pupil characteristics for these pupils.
[12] The Statistical Analysis Plan was developed and agreed upon with EEF in 2018, prior to the release of the revised 'EEF standard model.'

Secondary analysis

We analysed the ITT impact of the programmes using the equation (1) as described above. Only now, Y_{ijk} represents the secondary outcome measures for pupil i in classroom j in school k, described in more detail above. Specifically for FS Y_{ijk} represents the combined Year 6 and Year 7 KS2 writing results,[13] and for RWI Y_{ijk} represents both the KS1 writing percentage pass rate at the end of the second year and the Year 1 phonics screening check for both years of the programme aggregated (for Year 1 and Year 2 pupils). R_k and X_{ijk} remain as described in the primary outcome analysis section.

Analysis in the presence of non-compliance

As described in the randomisation section, 135 schools were randomised into the 2 groups, but we considered 131 schools for the ITT analysis. For our primary CACE analysis for RWI, we conducted analyses that consider whether the schools actually complied with the assignment.[14] Of the 131 schools, 68[15] schools reported receiving RWI or an aspect of RWI and 63[16] schools did not.[17] Our primary CACE analysis uses a broad definition of compliance by considering receipt of the intervention or an aspect of the intervention at the school level as compliance, rather than restricting compliance only to those schools which received the full RWI intervention. While this approach is imperfect, given the limitations of the available compliance data, this is sufficient. Our results are likely to be lower bounds because of the fuzziness of the compliance measure. RWI was delivered to all children, so pupils who have received the treatment in CACE analysis were all pupils from the schools that actually received RWI.

For FS, a number of treatment schools failed to deliver FS. Specifically, 23 schools (35%) delivered no FS programming at all,[18] 19 schools (29%) delivered FS to some pupils, but not all FS-eligible pupils, and 8 schools (12%) did not provide enough data for us to know whether or not they successfully delivered FS programming. Compliance with FS appears to have been mixed; however, it has not been possible to conduct a robust CACE analysis for FS. Due to a data matching issue, the FS CACE analysis was unable to be conducted as pupil compliance data could not be linked to outcome data from the NPD. Alternative approaches were considered but would have relied on strong and invalid assumptions, so analysis in the presence of non-compliance is not presented for FS. We are therefore unable to explore the relationship between exposure to FS and reading attainment as planned.

For RWI, we analysed the average impact of the programmes on compliers using the following two-stage least squares specification. In the first stage, depicted by equation (2), we estimate for treatment receipt based on treatment assignment. T_k is a variable that equals one if the school actually received the RWI programme. That is, T_k will equal one for the 69 schools that received treatment, whether they were allocated to the treatment group or not. As before, R_k is a variable that equals one if the pupil's school was assigned to the treatment and equals zero if not. X_{ijk} refers to the set of controls for the primary specification in equation (1).

[13] The FS pupils referred to as the 'Year 7 cohort' were in Year 7 when the evaluation ended but received the intervention and completed the outcome testing while in Year 6. The FS pupils referred to as the 'Year 6 cohort' were in Year 6 when the intervention ended. They began receiving the intervention while in Year 5 and completed outcome testing in Year 6.
[14] The 'primary outcome analysis' takes an ITT approach, meaning that every pupil included in the trial at the point of randomisation is included in the final analysis, even if they did not ultimately undertake the intervention. The 'primary outcome analysis' analyses differences between the outcomes of pupils randomised to the treatment group and those randomised to the control group. In contrast, the 'CACE analysis' analyses differences between the outcomes of pupils who ultimately took part in the programme in comparison to pupils who did not take part.
[15] Equal to the 66 treatment schools in the ITT analysis minus 11 treatment to non-treatment crossovers plus 13 control to treatment crossovers.
[16] Equal to the 65 control schools in the ITT analysis plus 11 treatment to non-treatment crossovers minus 13 control to treatment crossovers.
[17] Schools receiving 'an aspect' of the intervention include schools whose teachers received training from RMT within the last three years but did not purchase the programme materials, schools who received the programme in its entirety in the first year, but opted out in the second year, or schools who purchased materials from RMT or booked trainings even though they were assigned to the control group.
[18] These 23 schools include 11 treatment schools which delivered neither RWI or FS and 12 treatment schools which delivered RWI but not FS.

$$T_k = \alpha + \mu R_k + \rho X_{ijk} + \varepsilon_{ijk} \quad (2)$$

The first stage equation enables us to generate predicted values for actual treatment receipt, where the caret symbol indicates the value is estimated.

$$\hat{T}_k = \hat{\alpha} + \hat{\mu} R_k + \hat{\rho} X_{ijk} \quad (3)$$

We then used the predicted values of treatment receipt $\hat{T}_k$ from equation (3) in our estimate of the complier average causal effect through equation (4).

$$Y_{ijk} = \alpha' + \beta' \hat{T}_k + \theta' X_{ijk} + \varepsilon'_{ijk} \quad (4)$$

Note, actually conducting the two-stage procedure would produce incorrect standard errors. Statistical packages like Stata have built in modules like ivregress, which automatically correct the standard errors. The results of the two-stage least squares regression will provide us with an estimate of the CACE impact, the impact for those schools/pupils who complied with the programme. We expect that the CACE impact should be larger (more positive) than the ITT impact. Results of the CACE analysis can be found in the *Analysis in the presence of non-compliance results* section.

Missing data analysis

Since more than 5% of the total pupils had incomplete information, we first looked at the extent of the missing data for each variable to be included in the analysis and the patterns of missingness and consider multiple imputation as an alternative to listwise deletion if necessary.[19] We conducted t-test comparisons between pupils missing and not-missing data elements. We also explored a logistic model to explain the patterns of missingness using all available covariates. This descriptive analysis informed the choice of auxiliary variables to be included in the multiple imputation procedure. Specifically, we considered the missing group to be composed of FS pupils that were identified as eligible for the programme, for whom we have missing data, whereas the non-missing group was composed of pupils that were identified as eligible for the programme, for whom data are not missing. We did not analyse missingness of RWI pupil characteristics because we do not have data on pupil characteristics for most pupils in the RWI evaluation sample. Instead, we only analyse missingness of outcome data for these pupils. To see whether missingness led to an imbalance, we examined whether the non-missing variables were balanced between those who were in the missing group and the non-missing group. Appendix D reports the details of our missing data analysis. Since we found statistically significant differences (at the 5% significance level), we conducted multiple imputation with five imputations as a sensitivity analysis. We performed the multiple imputation using Stata's command 'mi impute mvn', which implements the multivariate normal regression method of imputation. We used Stata's default number of iterations for the burn-in period, which is 100, and assessed convergence of the imputed date by creating trace plots.

Subgroup analyses

We conducted subgroup analysis for the population of FSM pupils.[20] However, because of the small number of FSM-eligible pupils per school for FS, we did not have the power to detect a separate effect for FSM pupils, although the analysis was still conducted. We conducted the subgroup analysis in two ways. First, we estimated equation (1) only for the sample of FSM pupils. Second, we examined heterogeneous effects to see if there was a difference between FSM and non-FSM pupil with the equation:

$$Y_{ijk} = \alpha + \beta_1 R_k + \beta_2 R_k * F_{ijk} + \theta X_{ijk} + \delta F_{ijk} + \varepsilon_{ijk} \quad (5)$$

where F_{ijk} is an indicator for whether the pupil ever received FSM in the past six years. Here, the vector of controls X_{ijk} refers to the controls from the primary specification from equation (1), and for the FS analysis, the KS1 results act as a

[19] In general, 5% is the industry standard for the level of missing data which is acceptable and does not require imputation to correct for missingness.
[20] The FSM variable we use for the subgroup analyses will be the same that we describe above, namely whether the pupil was ever eligible for FSM (EVERFSM_6_p_ in the NPD).

baseline measure of the outcome variable. With the heterogeneous effects regression, β_1 represents the marginal effect of being part of the RWI or FS programmes, as compared to the control group, for pupils who do not receive FSM. β_2 is the marginal effect of being part of the RWI or FS programmes for pupils who do receive FSM. These heterogeneous effects analyses clarified if there is a difference in the programme effect by FSM status.

We also examined if heterogeneous effects exist along the dimension of gender with the following equation.

$$Y_{ijk} = \alpha + \beta_1' R_k + \beta_2' R_k * G_{ijk} + \theta' X_{ijk} + \delta' G_{ijk} + \varepsilon_{ijk} \quad (6)$$

where G_{ijk} is an indicator for whether the pupil is a girl. The vector of controls X_{ijk} is as defined for equation (5). With the heterogeneous effects regression, β_1' represents the marginal effect of being part of the RWI or FS programmes, as compared to the control group, for boys. β_2' is the marginal effect of being part of the RWI or FS programmes for girls. We also ran the results for female pupils only. These heterogeneous effects analyses clarified if there is a difference in the programme effect by gender.

Additional analyses and robustness checks

We conducted a sensitivity analysis comparing the primary results with results that include the two control schools that had received the intervention training within the last three years and did not disclose that information in their registration materials. For the sensitivity analysis, we analysed the ITT impact of the programmes using equation (1) as described above but with 133 schools instead of 131.

Estimation of effect sizes

We calculated effect sizes using the unadjusted pooled variance of the outcome. The numerator was the coefficient on the school treatment indicator (a regression-adjusted impact estimate), and the denominator was calculated using the observations included in the regression according to the following formula (EEF, 2018):

$$s^* = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Specifically, we calculate the Hedges' g effect size using the following formula:

$$\text{Hedges}'\ g = \frac{\mu_1 - \mu_2}{s^*}$$

where $\mu_1 - \mu_2$ is the difference in means between the treatment group and control group, respectively.

Estimation of ICC

We calculated ICCs using the loneway command in Stata.

Implementation and process evaluation

Table 5: IPE methods overview

Research methods	Data collection methods	Participants/ data sources	Data analysis methods	Research questions addressed	Implementation / logic model relevance
Fidelity to teaching survey	Self-administered survey	Reading leaders, teachers	Summary statistics	RQs 2–5	To get self-reported information on the implementation fidelity of the RWI and FS programmes
Key informant interviews	Recorded interview	Headteachers, reading leaders, trainers, RMT staff	Deductive coding, analysis and summation	RQs 2–6	To understand the implementation of reading leader and teacher training, reading leader support and continuous professional development
Focus group discussions	Recorded interview	Teachers, teaching assistants	Deductive coding, analysis and summation	RQs 2–8	To understand the implementation of teacher training, reading leader support, continuous professional development and teacher practices and behaviours in the classroom
Ethnographic observations	Fieldnotes	Teacher training session RWI lessons	Deductive coding, analysis and summation	RQs 2–3, 8	To understand teacher practices and behaviours in the classroom

AIR conducted a relatively light-touch process snapshot (including seven key informant interviews, two focus group discussions and 14 rapid ethnographic observations in two programme schools) to check that the programmes were delivered as intended, to understand the contrast from usual practice and to identify any barriers to implementation and necessary conditions for success.

For this process snapshot, we analysed RMT survey data and collected qualitative data from treatment schools.[21] First, we analysed the results of a survey conducted by RMT trainers on teaching and training fidelity in treatment schools that focuses on factors identified by programme implementers as the most crucial to the success of the programme. Second, we conducted qualitative interviews with RMT staff and school administrators and teachers in two treatment schools and one control school to triangulate the information obtained from the learners.

AIR contributed to the design and content of the various questionnaires that were implemented by RMT. To reduce concerns associated with response bias, whereby teachers answer with what they think the researchers want, we used questions about actual day-to-day teaching practices. These questionnaires as well as the qualitative interviews helped us address the following research questions:

1. How are the programmes being implemented?
2. Are the teachers using the programmes as envisioned by the developers?
3. Are teachers receiving support from the reading leader?
4. How does training quality help support the quality of the reading leader to implement the programme with fidelity?
5. How contingent is the success of the programme on this continuous professional development?
6. How has the programme impacted teacher knowledge on how to teach reading?
7. Does the programme change teacher practices and behaviours in the classroom?

[21] Treatment schools here refer to schools randomly assigned to receive the intervention and were, subsequently, trained by RMT on RWI and FS and implemented the programmes in their school.

Qualitative process evaluation

AIR's qualitative approach to the process evaluation sought to understand how the RWI and FS programmes were implemented, including barriers to delivery and necessary conditions for success. To gain knowledge on the background of the programme and understand training standards and practices, AIR conducted three key informant interviews with RMT programme staff and trainers and ethnographic observation of one phonics teacher training at RMT headquarters in London. To understand perceptions of programme delivery, structures of training and professional development support, administrator and teacher practices, and identify barriers and facilitators to implementation, AIR conducted four key informant interviews, two focus groups and 14 rapid ethnographic observations of RWI and FS lessons in two treatment schools. In addition, AIR analysed information from MoUs to understand existing phonics and reading programmes in control schools, and conducted one key informant interview in a control school. AIR used data from interviews, focus groups and observations to contextualise and triangulate information obtained from teacher surveys as well as from the learners.

Qualitative sampling

For the qualitative portion of this evaluation, we visited two programme schools and one control school. For programme schools, we purposively sampled schools based on performance and geographic distribution. We selected one high-performing and one low-performing school, based on Ofsted and phonics screening results. We sampled one school located in northern England and one located near London, to maximise regional variation and facilitate data collection logistics. For the control school, we sampled a relatively low-performing school located near London to facilitate logistics and deepen our understanding of the struggles of teaching literacy and reading in our control sample.

At each programme school, we conducted key two informant interviews with school administrators and reading leaders, as well as one focus group discussion with teachers and teaching assistants involved in implementing the RWI programme. Furthermore, we conducted brief observations of RWI and FS sessions in each school (total 14 observations).

In the control school, we conducted key informant interviews with two administrators involved in coordinating the school's reading programme. Due to busy school schedules, literacy teachers were not available for focus group discussions during the control school visit.

In addition, to understand implementation from the perspective of programme staff, we conducted key informant interviews with two trainers and the director of RMT as well as observation of one teacher training session at RMT headquarters in London.

Addressing potential biases in qualitative data collection

With any form of qualitative or quantitative research, minimising researcher bias is of critical importance to ensuring the validity of findings. Bias management is essential when conducting interviews and FGDs, to ensure the collection of high-quality, rigorous, and unbiased data (Chenail, 2011). To address potential bias in the collection, analysis, and reporting of qualitative data, AIR designed and pre-tested research protocols to ensure that questions are open-ended, and not leading or biased (White & Phillips, 2012). During data collection, AIR conducted member checks with interviewees when appropriate throughout the research process (Creswell & Miller, 2000). Further, the AIR study team also gathered evidence for potential, external influencing factors to mitigate the tendency for evaluators to see a relationship between the programme and outcomes where none exists. This process evaluation has also been designed such that data collected from KIIs, FGDs and observations complement and help to triangulate findings from the fidelity to teaching survey.

Qualitative analysis

Analysis of the qualitative data followed a rigorous and iterative process of coding, analysis and summation. AIR researchers created a deductive coding structure based on the interview and focus group protocols, which served as a preliminary guide for data analysis. AIR then reviewed interview and focus group transcripts, as well as observation protocols, to identify additional themes for the coding structure. By coding interview and focus group transcripts (using the qualitative data analysis software NVivo), the AIR study team identified patterns of response that emerge across categories and individuals (Maxwell, 2005). Furthermore, transcripts were informed by the larger understanding of the

conceptual framework of the RWI programme. After our analysis of the data, we synthesised and summarised the findings and examine these findings compared with other data sources.

Costs

Cost information was collected through documentation by the RMT team, headteachers and school administrators from intervention schools. We then estimate the costs of the programme for the average beneficiary and divide these costs by the estimated gain in outcome derived from the impact analysis to serve as the cost-effectiveness measure of the programmes.

RMT followed a train-the-trainers method. This approach involved two levels of start-up cost: one for the trainers and one for the teachers. We make the assumption that teachers stay at a school for at least three years and further that if teachers are trained once on the RWI and FS programmes, the fixed costs can be spread across three annual cohorts of teachers.

Timeline

Table 6 presents the key dates related to the delivery of the RWI and FS programmes for this evaluation, including the training of the trainers and the recruitment of study participants and treatment assignment. The table further summarises when the final pupil assessments and teacher surveys were administered.

Table 6: Timeline

Dates	Activity	Staff responsible / leading
April 2016	AIR randomised first batch of schools	AIR
May 2016	AIR randomised second batch of schools	AIR
June 2016	AIR randomised final batch of schools	AIR
Summer 2016	RMT delivers training of the trainers	RMT
September 2016	Schools begin implementing RWI and FS	Schools
September 2016	QUB submitted initial request of NPD data	QUB
May 2017	FS Year 7[22] cohort sat KS2 assessments	Schools
April 2018	QUB administered teacher survey	QUB

[22] The year group refers to the year pupils were in at the end of the evaluation, in summer 2018. The FS pupils referred to as the 'Year 7 cohort' were in Year 7 when the evaluation ended, but received the intervention and completed the outcome testing while in Year 6. The FS pupils referred to as the 'Year 6 cohort' were in Year 6 when the intervention ended. They began receiving the intervention while in Year 5 and completed outcome testing in Year 6.

Dates	Activity	Staff responsible / leading
May 2018	FS Year 6 cohort sat KS2 assessments	Schools
May–July 2018	QUB administered NGRT	QUB
June 2018	AIR conducted focus group discussion and key Informant Interviews	AIR
July 2018	School trials end	Schools
November 2018	QUB submitted request of NPD data	QUB
September 2019–January 2021	AIR conducted analyses and drafted report (process delayed due to data access issues)	AIR
February 2021	Draft report submitted	AIR

Impact evaluation

Participant flow including losses and exclusions

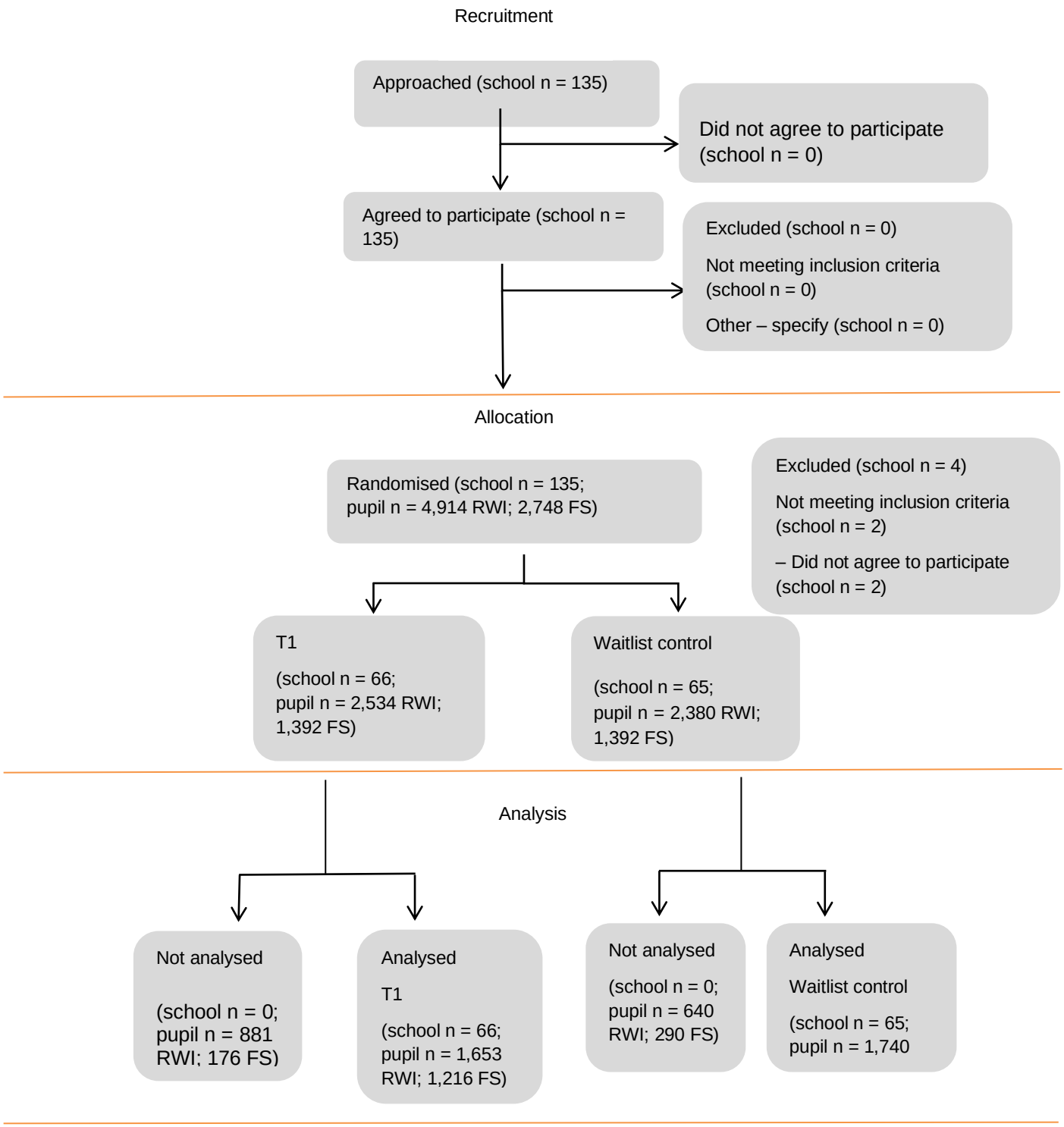
Figure 3 shows the participant flow diagram that documents experiences of the schools in the trial. 135 schools were randomised into the two groups, with 67 assigned to the control and 68 assigned to the treatment. However, two of the schools assigned to the treatment withdrew before they were informed of the results of the randomisation. Thus, we excluded these two schools from all analyses. Additionally, after randomisation results were shared, it was realised that two control schools had received the intervention training within the last three years and did not disclose that information in their registration materials. Had this information been properly disclosed they would not have been eligible. For the primary ITT analysis, these four schools were excluded. Thus, for the ITT analysis, we considered 131 schools, 65 in the control group and 66 in the treatment group, which corresponds to the 135 randomised schools minus the two treatment that withdrew prior to being informed of their randomisation results and the two control schools who were ineligible. We conducted a sensitivity analysis comparing the primary results with results that include the two control schools who were ineligible.

A number of treatment schools did not offer the intervention and a number of controls schools offered varying degrees of the intervention. Specifically, 10 schools (15%) that were randomised to intervention and informed of their results did not deliver the RWI or FS intervention, although two of these schools received the initial training. An additional school received the programme in its entirety for the first year but opted out for the second year.

Additionally, five control schools withdrew from the trial. All of these except for one school, booked intervention training and/or purchased RWI materials from RMT after trial randomisation. Four additional control schools also attended regional training administered by RMT after trial randomisation but did not withdraw from the study. Five additional control schools purchased RWI resources after the trial began but did not withdraw from the study. To reduce the possibility of similar instances occurring, all control schools were notified of the importance of maintaining their control status for the research and of the incentive that will be provided at the end of the trial.

The statistical power at each stage of the study (protocol, randomisation, analysis) is given in Table 7 for both RWI and FS, respectively. The power calculations in the protocol column reflect the MDES based on assumptions made during the design phase. Those in the randomisation column reflect slight changes based on revised sampling. The calculations in the final column reflect the final model used for analysing the primary outcome for RWI and FS and the final sample sizes for those models. We only include the proportion of variance explained by the prior KS1 test scores for FS power calculations as there are no relevant baseline scores to use for the RWI analysis. The power calculations for RWI are based on the sample sizes for Year 2 pupils at the end of the intervention period as those are the only ones who took the NGRT assessment, the primary outcome for the RWI ITT analysis.

Figure 3: Participant flow diagram (2 arms)[23]



[23] Four schools dropped out of our study after randomisation. However, this was not biased dropout as two of these schools dropped out prior to learning the results of the randomisation process, and the other two schools received intervention training within the last three years without disclosing this information so would not have been eligible for the trial.

Table 7: Minimum detectable effect size at different stages

		Protocol				Randomisation				Analysis			
		RWI		FS		RWI		FS		RWI		FS	
		Overall	FSM	Overall	FSM	Overall	FSM	Overall	FSM	Overall	FSM	Overall	FSM
MDES		0.184	0.222	0.215	0.267	0.176	0.222	0.193	0.205	0.207	0.241	0.203	0.232
Pre-test/post-test correlations	Level 1 (pupil)	0.00	0.00	0.53	0.53	0.00	0.00	0.53	0.53	0.00	0.00	0.18	0.18
Intra-cluster correlations (ICCs)	Level 2 (school)	0.103	0.103	0.133	0.133	0.103	0.103	0.133	0.133	0.143	0.143	0.133	0.133
Alpha		0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
Power		0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
One-sided or two-sided?		two	two	two	two	two	two	two	two	two	two	two	two
Average cluster size		37	11	10	3	38	9	21	10	26	9	20	8
Number of schools	Intervention	60	60	60	60	66	66	66	66	66	66	66	66
	Control	60	60	60	60	65	65	65	65	65	65	65	65
	Total	120	120	120	120	131	131	131	131	131	131	131	131
Number of pupils	Intervention	2,220	660	600	180	2,534	590	1,392	677	1,653	735	1,354	406
	Control	2,220	660	600	180	2,380	590	1,392	677	1,740	500	1,219	585
	Total:	4,440	1,320	1,200	360	4,914	1,180	2,748	1,354	3,393	1,235	2,573	991

Attrition

For RWI and FS, we examine the attrition of pupils that were randomised for the programme but did not remain in our analytic sample for our primary outcome analysis for any reason. Specifically, we consider the attrition group to be composed of eligible, randomised pupils for whom we do not have primary outcome data. For RWI, this group includes pupils in the Year 2 cohort for whom we do not have NGRT scores, while for FS, this group includes pupils in both cohorts for whom we have KS1 results but do not have KS2 reading results. Overall, we have NGRT scores for about 70% of the RWI pupils and have both KS1 and KS2 reading results for approximately 90% of the FS pupils included at the stage of randomisation. Some pupils were lost due to dropping out of school, moving to a different school, absence on the day of the national examinations, or a variety of other reasons. Tables 8a and 8b present the number and proportion of pupils in the attrition group by intervention group.

Table 8a: Pupil level attrition from the RWI trial (primary outcome)

		Intervention	Control	Total
Number of pupils	Randomised	2,534	2,380	4,914
	Analysed	1,653	1,740	3,393
Pupil attrition (from randomisation to analysis)	Number	881	640	1,521
	Percentage	34.77	26.89	30.95

For the RWI primary analysis, we examine NGRT scores for the Year 2 cohort only. Accordingly, they are the only group for which we assess attrition in line with EEF's definitions and ratings. The proportion of pupils in the attrition group for RWI is high in both groups, but higher in the intervention group compared to the control group and is approximately 31% overall. This attrition likely comes from pupils missing the testing date, teachers stopping administration of the NGRT due to perceived difficulty, pupils moving school since randomisation, as well as other reasons. We were not able to conduct further analyses to examine whether attrition led to an imbalance in our Year 2 RWI sample as we do not have pupil-level characteristics for this group.

Table 8b: Pupil level attrition from the FS trial (primary outcome)

		Intervention	Control	Total
Number of pupils	Randomised	1,392	1,392	2,784
	Analysed	1,216	1,102	2,318
Pupil attrition (from randomisation to analysis)	Number	176	290	466
	Percentage	12.64	20.83	16.74

The proportion of pupils in the attrition group is slightly higher in the control group compared to the intervention group, so we further examine whether this attrition led to an imbalance in our FS sample. We find that FS pupils in the attrition group are more likely to be FSM (p-value = 0.01) and have lower KS1 reading scores (p-value = 0.00). Understanding that FSM pupils are more likely to change school both within and across school years, these results are unsurprising. Table 9 presents the comparison of FS pupil characteristics by attrition status.

Table 9: Pupil level attrition from the FS trial (primary outcome)

Variable	Attritors		Panel				
	Mean	N	Mean	N	Diff	SE	p-value of diff.
Female	0.37	211	0.41	2,573	0.04	0.03	0.201
Year 6[24]	0.61	211	0.49	2,573	−0.12	0.08	0.138
Year 7	0.39	211	0.51	2,573	0.12	0.08	0.138
Ethnicity							
White British	0.72	211	0.76	2,573	0.07	0.05	0.126
Other White	0.11	211	0.05	2,573	−0.07	0.04	0.084
Mixed	0.02	211	0.03	2,573	0.01	0.01	0.374
Other Asian	0.02	211	0.06	2,573	0.02	0.02	0.149
Black African	0.02	211	0.02	2,573	0.00	0.09	0.953
Other ethnicity	0.02	211	0.02	2,573	0.00	0.01	0.560
Other Characteristics							
ever FSM	0.62	211	0.50	2,573	−0.10	0.04	0.010
EAL	0.11	211	0.13	2,573	0.00	0.03	0.910
KS1 reading points	9.66	211	13.00	2,573	3.38	0.74	0.000

Pupil and school characteristics

Schools in our study were assigned to intervention or control status within strata based on their Ofsted scores and KS1 reading score quartiles. Additional school-level variables were obtained from the NPD. For each characteristic, we found no statistically significant difference between intervention and control schools (Table 10).[25]

[24] The FS pupils referred to as the 'Year 7 cohort' were in Year 7 when the evaluation ended but received the intervention and completed the outcome testing while in Year 6. The FS pupils referred to as the 'Year 6 cohort' were in Year 6 when the intervention ended. They began receiving the intervention while in Year 5 and completed outcome testing in Year 6.
[25] Significance was tested using t-tests of balance (Stata test command).

Table 10: Baseline school-level characteristics of groups as randomised

School-level (categorical)	National-level mean	Intervention group		Control group	
		n/N (missing)	Count (%)	n/N (missing)	Count (%)
Ofsted score					
1		<10/66 (0)	<10 (<15.15%)	14/67 (0)	14 (20.90%)
2		47/66 (0)	47 (71.21%)	43/67 (0)	43 (64.18%)
3		10/66 (0)	10 (15.15%)	<10/67 (0)	<10 (<15.15%)
4		<10/66 (0)	<10 (<15.15%)	<10/67 (0)	<10 (<15.15%)
KS1 quartile					
1		17/66 (0)	17 (25.876%)	14/67 (0)	14 (20.90%)
2		15/66 (0)	15 (22.73%)	16/67 (0)	16 (23.88%)
3		17/66 (0)	17 (25.76%)	16/67 (0)	16 (23.88%)
4		17/66 (0)	17 (25.76%)	21/67 (0)	21 (31.24%)
Urban/rural school type					
Urban		48/66 (0)	48 (72.73%)	48/67 (0)	48 (71.64%)
Rural		18/66 (0)	18 (27.27%)	19/67 (0)	19 (28.36%)
Percent of pupils FSM-eligible					
Low (less than 33%)		39/66 (0)	39 (59.09%)	43/67 (0)	43 (64.18%)
Medium (33% to 66%)		20/66 (0)	20 (30.30%)	20/67 (0)	20 (29.85%)
High (greater than 66%)		7/66 (0)	7 (10.61%)	4/67 (0)	4 (5.97%)
School-level (continuous)					
KS1 L2 reading	0.75	66/66 (0)	0.88 (0.10)	67/67 (0)	0.89 (0.08)

At the pupil level, we find a few statistically significant differences between intervention and control groups for RWI (Table 11) and FS (Table 12). For RWI, pupils randomised to the intervention group are more likely to be Other White

(5%) and Mixed ethnicity (4%) compared to pupils in the control group (3% and 3%, respectively).[26] Intervention pupils are also more likely to be EAL (13%) compared to those in the control group (almost 8%) which could result in some underestimation of the impact of RWI on reading comprehension.[27]

Table 11: Baseline pupil-level characteristics of groups as randomised for RWI

Pupil-level (categorical)	National-level mean	Intervention group		Control group	
		n/N (missing)	Count (%)	n/N (missing)	Count (%)
Year group[28]					
Year 1		1,732/6,703 (0)	1,732 (25.84%	2,004/6,735 (0)	2,004(29.76%)
Year 2		2,534/6,703 (0)	2,534 (37.80%)	2,380/6,735 (0)	2,380 (35.34%)
Year 3		2,437/6,703 (0)	2,437 (36.36%)	2,351/6,735 (0)	2,351(34.91%)
Gender					
Female		2,473/6,703 (1,732)	2,473 (36.89%)	2,138/6,735 (2,061)	2,138 (31.74%)
Male		2,498/6,703 (1,732)	2,498 (37.27%)	2,536/6,735 (2,061)	2,536 (37.65%)
Ethnicity					
White British		3,310/6,703 (2,058)	3,310 (49.38%)	3,672/6,735 (2,604)	3,672 (54.52%)
Other White		339/6,703 (2,058)	339 (5.06%)	189/6,735 (2,604)	189 (2.81%)
Mixed		284/6,703 (2,058)	284 (4.24%)	179/6,735 (2,604)	179 (2.66%)
Other Asian		393/6,703 (2,058)	393 (5.86%)	205/6,735 (2,604)	205 (3.04%)
Black African		126/6,703 (2,058)	126 (1.88%)	111/6,735 (2,604)	111 (1.65%)
Black Caribbean		23/6,703 (2,058)	23 (0.34%)	27/6,735 (2,604)	27 (0.40%)

[26] T-test p-values of 0.018 and 0.038, respectively, assuming equal variances.
[26] T-test p-value of 0.044.
[28] The year group refers to the year pupils were in at the end of the evaluation, in summer 2018. The RWI pupils referred to as the Year 1 cohort were in Year 1 when the evaluation ended but began receiving the intervention while in Reception and completed outcome testing in Year 1. The RWI pupils referred to as the Year 2 cohort were in Year 2 when the evaluation ended but began receiving the intervention while in Year 1 and completed outcome testing in both Year 1 and Year 2. The RWI pupils referred to as the Year 3 cohort were in Year 3 when the evaluation ended but received the intervention and completed outcome testing while in Year 2.

Other Black		38/6,703 (2,058)	28 (0.57%)	34/6,735 (2,604)	34 (0.50%)	
Bangladeshi		25/6,703 (2,058)	25 (0.37%)	32/6,735 (2,604)	32 (0.47%)	
Chinese		20/6,703 (2,058)	20 (0.30%)	24/6,735 (2,604)	24 (0.36%)	
Other Characteristics						
Other ethnicity		87/6,703 (2,058)	87 (1.30%)	57/6,735 (2,604)	57 (0.84%)	
ever FSM		1,563/6,703 (1,732)	1,563 (23.32%)	1,175/6,735 (2.059)	1,175 (23.58%)	
EAL		894/6,703 (1,737)	894 (13.31%)	520/6,735 (2,070)	520 (7.72%)	

For FS pupils, we find that pupils randomised to the intervention group are more likely to be Other White (8%) compared to pupils in the control group (3%).[29] However, control pupils are more likely to be White British (82%) compared to pupils in the intervention group (69%).[30] Intervention pupils are also more likely to be EAL (17%) compared to control pupils (8%).[31] We find the groups scored similarly on the KS1 reading exam; however, overall, pupils in our sample scored lower than the national average on KS2 reading by 10 points, on average (23.4 points in our sample compared to 33 points nationally). Given that the FS programme is specifically targeted to pupils for remedial literacy instruction, our sample scoring lower than the national average on the KS2 reading exams is unsurprising. Even so, the results of our FS evaluation are only generalisable to other pupils who would qualify for the remedial programming, rather than the broader Year 6 and Year 7 cohorts, nationally.

[29] t-test p-value of 0.006.
[30] t-test p-value of 0.040.
[31] t-test p-value of 0.020.

Table 12: Baseline pupil-level characteristics of groups as randomised for FS

Pupil-level (categorical)	National-level mean	Intervention group		Control group	
		n/N (missing)	Count (%)	n/N (missing)	Count (%)
Year 6[32]		670/1,354 (0)	670 (49.48%)	638/1,277 (0)	638 (49.96%)
Year 7		684/1,354 (0)	684 (50.52%)	639/1,277 (0)	639 (50.04%)
Female		551/1,354 (0)	551 (40.69%)	525/1,277 (0)	525 (41.11%)
Male		803/1,354 (0)	803 (59.31%)	752/1,277 (0)	752 (58.89%)
Ethnicity					
White British		940/1,354 (65)	940 (69.42%)	1,047/1,277 (67)	1,047 (81.99%)
Other White		105/1,354 (65)	105 (7.75%)	42/1,277 (67)	42 (3.29%)
Mixed		44/1,354 (65)	44 (3.25%)	42/1,277 (67)	42 (3.29%)
Other Asian		104/1,354 (65)	104 (7.68%)	36/1,277 (67)	36 (2.82%)
Black African		33/1,354 (65)	33 (2.44%)	14/1,277 (67)	14 (1.10%)
Black Caribbean		12/1,354 (65)	12 (0.89%)	–	–
Other Black		14/1,354 (65)	14 (1.03%)	–	–
Bangladeshi		10/1,354 (65)	10 (0.74%)	–	–
Other ethnicity		27/1,354 (0)	27 (1.98%)	29/1,277 (67)	29 (2.26%)
ever FSM		711/1,354 (0)	711 (52.51%)	630/1,277 (0)	630 (49.33%)

[32] The FS pupils referred to as the 'Year 7 cohort' were in Year 7 when the evaluation ended but received the intervention and completed the outcome testing while in Year 6. The FS pupils referred to as the 'Year 6 cohort' were in Year 6 when the intervention ended. They began receiving the intervention while in Year 5 and completed outcome testing in Year 6.

EAL		236/1,354 (0)	236 (17.43%)	105/1,277 (0)	105 (8.22%)	
Pupil-level (continuous)		**n/N (missing)**	**Mean (SD)**	**n/N (missing)**	**Mean (SD)**	**Effect size**
KS1 reading points	N/A	1,349/1,354 (5)	12.69 (3.65)	1,271/1,277 (6)	12.66 (3.51)	−0.007
KS1 z-scores	N/A	1,349/1,354 (5)	−0.86 (0.93)	1,271/1,277 (6)	−0.87 (0.89)	−0.007

Outcomes and analysis

Primary analysis

We find small (in terms of effect size) insignificant impacts of the RWI programme on the primary outcome (NGRT SAS). The mean score for pupils in the intervention schools was 97.04 points compared to 96.84 points, on average, among the control group, though not statistically significant. This difference translates to an effect size of +0.05 or roughly one month's progress. Table 13 presents the results of the primary RWI analysis while details of the effect-size calculations can be found in Table C1 in Appendix C. Sample analysis code is provided in Appendix E.

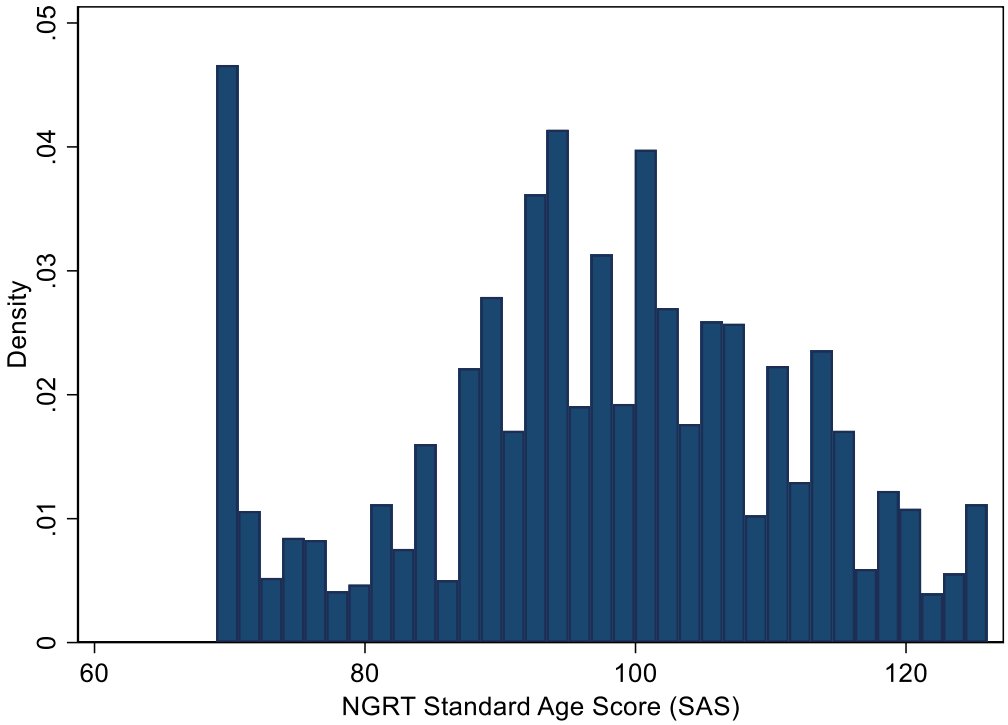
Table 13: Primary analysis for RWI[33]

	Unadjusted means				**Effect size**		
	Intervention group		**Control group**				
Outcome	**n (missing)**	**Mean (95% CI)**	**n (missing)**	**Mean (95% CI)**	**Total n (intervention; control)**	**Hedges' g (95% CI)**	**p-value**
NGRT – SAS	1,653 (881)	97.04 (96.33–97.75)	1,740 (640)	96.84 (96.16–97.51)	3,393 (1653; 1740)	0.05 (−0.02, 0.12)	0.528

Figure 4 presents the distribution of NGRT scores. Based on this distribution we observe clear floor effects with a clustering of scores around the bottom, suggesting the difficulty of the assessment may have been too high for many pupils. The presence of such a floor effect limits our ability to accurately measure the true average reading scores and abilities of pupils, as we are unable to ascertain nuances at the lower threshold. Further, this distribution supports the notion that the NGRT was a more difficult assessment than typical grade-appropriate reading assessments for Year 2 pupils including the KS1 exams. As mentioned in the attrition section, there were some teachers who raised concerns about the difficulty of the assessment, and subsequently stopped administering the test to their pupils for this reason.

Figure 4: Histogram of NGRT SAS

[33] Given that pupils in the Year 2 cohort were in Reception at randomisation, it is possible there are pupils for whom we collected UPNs in Reception, since they were at the school at the time of trial randomisation, but who since transferred to a different school by Year 2 such that we do not have NGRT scores for them in our sample. These pupils may appear as missing in our data since they would have transferred schools, not received RWI, and not have sat the NGRT but the randomisation for the trial was at the school level rather than the pupil. In other words, pupils who moved before Year 1 should not count against missing values but may have in this case. However, there is only a discrepancy of 21 pupils for which this error may have been made such that we are not concerned with it biasing our results.



For FS, we find larger, negative effects of allocation to the FS treatment group on pupils' reading scores, though we only find evidence of statistical significance for eligible pupils in Year 7 at the end of the trial, or those who received the intervention for only one year (p-value = 0.015). The mean KS2 reading score for Year 7 pupils in intervention schools was 20.24 compared to an average of 23.02 points for eligible Year 7 pupils in control schools. Accounting for pupils' KS1 reading scores, the estimated impact of FS on reading scores was an effect size of −0.22.[34]

Similarly, the mean KS2 reading score was lower for Year 6 pupils in intervention schools (24.74 points) compared to those in control schools (25.55 points) (pupils who had the intervention for two years). The effect size among this cohort was −0.03 but not statistically significant (p-value = 0.699). When combining scores for both cohorts, the results follow a similar trend with the mean scores for pupils in the intervention schools lower than those for pupils in control schools (22.44 points compared to 24.26 points). We again estimate a negative effect size (−0.13), which is not statistically significant (p-value = 0.090).

Table 14 presents the full results of the primary analysis for FS. We provide details for the calculation of effect sizes in Table C2 in Appendix C along with sample analysis code in Appendix E.

[34] The KS1 reading scores from 2016 were based on the old examination system whilst the KS2 examinations delivered to pupils in 2018 were based on the higher expectations of the new National Curriculum. However, these different national examination systems and standards would affect the control and treatment groups similarly so should not cause the impact of FS to spuriously appear lower.

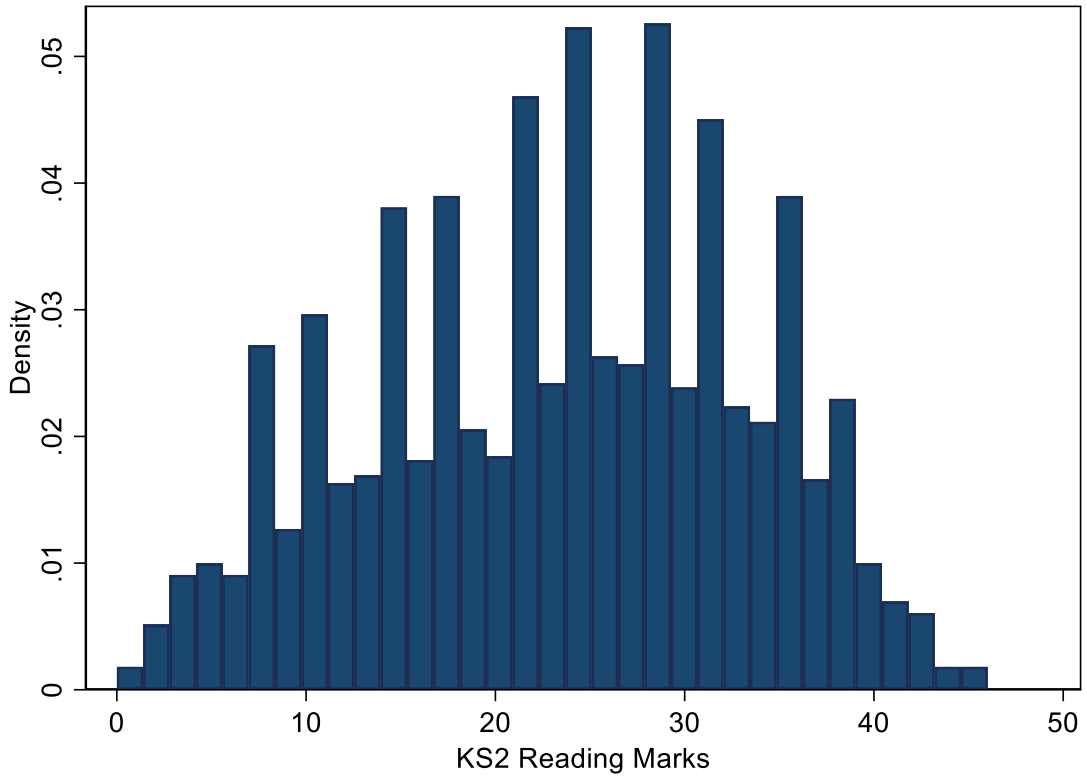
Table 14: Primary analysis for FS

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	**n (missing)**	**Mean (95% CI)**	**n (missing)**	**Mean (95% CI)**	**Total n (intervention; control)**	**Hedges' g (95% CI)**	**p-value**
KS2 reading marks – Year 7[35]	637 (47)	20.24 (19.62–21.06)	561 (54)	23.02 (22.26–23.77)	1,198 (637; 561)	−0.22 (−0.33, −0.11)	0.015
KS2 reading marks – Year 6	579 (91)	24.74 (23.91–25.58)	541 (63)	25.55 (24.73–26.38)	1,120 (579; 541)	−0.03 (−0.15, 0.09)	0.699
KS2 reading marks – combined	1,216 (138)	22.44 (21.87–23.00)	1,102 (117)	24.26 (23.70–24.83)	2,318 (1,216; 1,102)	−0.13 (−0.21, −0.05)	0.090

Figure 5 shows the distribution of combined KS2 reading marks for pupils in the Year 6 and Year 7 evaluation cohorts (i.e., scores for KS2 reading assessments sat by the Year 7 cohort while in Year 6 in 2017 and by the Year 6 cohort in 2018). Unlike with the distribution of NGRT scores, the KS2 reading marks exhibit a fairly normal distribution with no evident ceiling or floor effects enabling a more reliable estimate of pupil abilities.

[35] The FS pupils referred to as the 'Year 7 cohort' were in Year 7 when the evaluation ended but received the intervention and completed the outcome testing while in Year 6. The FS pupils referred to as the 'Year 6 cohort' were in Year 6 when the intervention ended. They began receiving the intervention while in Year 5 and completed outcome testing in Year 6.

Figure 5: Histogram of KS2 reading marks – combined



Secondary analysis

Using the same primary model specification, we examined effects of RWI on secondary outcomes (proportion passing KS1 writing assessment, phonics screening scores and proportion passing phonics screening assessment). Table 15 presents full results of this secondary analysis. We find consistently small, negative effect size impacts of the programme on the proportion of pupils passing KS1 writing assessment, though none of these results are statistically significant. We did find statistically significant effects with positive effect sizes on pupils' phonics screening scores and passage for Year 1 pupils only.

The proportion of pupils in Years 2 and 3 combined that passed the KS1 writing exam was 65% in the intervention schools compared to 68% in control schools. This estimated impact of RWI on the proportion passing the KS1 writing assessment was an effect size of −0.04. We found qualitatively similar effect sizes when examining the proportion passing the writing assessment individually by cohort: −0.04 for pupils in Year 2 and −0.06 for pupils in Year 3.

We found more promising results when looking at scores from the Year 1 phonics screening assessment. The mean phonics screening score across cohorts in the intervention schools was 34.11 points compared to 33.62 points in control schools. This difference represented an effect size of +0.08 and was not significant (p-value = 0.085). When looking at phonics screening scores by individual cohorts, we found moderately large, positive effects on Year 1 pupils (effect size = +0.18; p-value = 0.002). The mean score differentials were not as large for pupils in the Year 2 cohort: 33.38 points for the intervention schools compared to 33.39 points for the control schools. The effect size was +0.01 and was not statistically significant (p-value = 0.745). Combining the cohorts together, we found an overall effect of RWI on phonics screening scores of +0.08, equivalent to one month's additional progress. This estimate was not significant (p-value = 0.085).

Results on the proportion of pupils passing the phonics screener mirror those for the raw scores with positive, significant effects found for the Year 1 cohort (effect size = +0.15, p-value = 0.006), and smaller, albeit not significant, effects found for the combined cohorts (effect size = +0.05, p-value = 0.146).

Table 15: Secondary analysis for RWI

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	**n (missing)**	**Mean (95% CI)**	**n (missing)**	**Mean (95% CI)**	**Total n (intervention; control)**	**Hedges' g (95% CI)**	**p–value**
KS1 writing passed – Year 3[36]	2,403 (34)	0.63 (0.61–0.65)	2,320 (31)	0.66 (0.64–0.68)	4,723 (2,403; 2,320)	−0.06 (−0.12, −0.01)	0.254
KS1 writing passed – Year 2	2,496 (38)	0.67 (0.65–0.69)	2,343 (37)	0.70 (0.68–0.72)	4,839 (2,496; 2,343)	−0.04 (−0.10, 0.01)	0.265
KS1 writing passed – combined	4,899 (72)	0.65 (6.63–0.66)	4,663 (68)	0.68 (0.67–0.69)	9,562 (4,899; 4,663)	−0.04 (−0.08, −0.002)	0.231
Phonics screening score – Year 2	2,492 (42)	33.38 (33.04–33.73)	2,355 (25)	33.39 (33.04–33.74)	4,847 (2,492; 2,355)	0.01 (−0.04, 0.07)	0.745
Phonics screening score – Year 1	1,710 (22)	35.16 (34.83–35.49)	1,989 (0)	33.89 (33.52–34.26)	3,699 (1,710; 1,989)	0.18 (0.11, 0.24)	0.002
Phonics screening score – combined	4,202 (64)	34.11 (33.86–34.35)	4,344 (25)	33.62 (33.36–33.87)	8,546 (4,202; 4,344)	0.08 (0.04, 0.12)	0.085
Passed phonics – Year 2	2,492 (42)	0.82 (0.80–0.83)	2,355 (25)	0.82 (0.80–0.83)	4,847 (2,492; 2,355)	0.00 (−0.06, 0.06)	0.783
Passed phonics – Year 1	1,710 (22)	0.89 (0.88–0.91)	1,989 (0)	0.85 (0.83–0.87)	3,699 (1,710; 1,989)	0.15 (0.08, 0.21)	0.006
Passed phonics – combined	4,202 (64)	0.85 (0.84–0.86)	4,344 (25)	0.83 (0.82–0.84)	8,546 (4,202; 4,344)	0.05 (0.01, 0.10)	0.146

The secondary analysis for FS examined impacts on KS2 writing scores for both Year 6 and Year 7 pupils combined. We found a small, positive effect size estimate of +0.02, though this was not statistically significant (p-value = 0.815).

[36] The year group refers to the year pupils were in at the end of the evaluation, in summer 2018. The RWI pupils referred to as the Year 1 cohort were in Year 1 when the evaluation ended but began receiving the intervention while in Reception and completed outcome testing in Year 1. The RWI pupils referred to as the Year 2 cohort were in Year 2 when the evaluation ended, but began receiving the intervention while in Year 1 and completed outcome testing in both Year 1 and Year 2. The RWI pupils referred to as the Year 3 cohort were in Year 3 when the evaluation ended but received the intervention and completed outcome testing while in Year 2.

The average writing score for pupils in the intervention schools was 94.19 points compared to 94.18 points for pupils in control schools. Full results are presented in Table 16.

Table 16: Secondary analysis for FS

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	**n (missing)**	**Mean (95% CI)**	**n (missing)**	**Mean (95% CI)**	**Total n (intervention; control)**	**Hedges' g (95% CI)**	**p-value**
KS2 writing score – combined	1,324 (30)	94.19 (93.68–94.70)	1,188 (31)	94.18 (93.65–94.70)	2,512 (1,324; 1,188)	0.02 (−0.06, 0.10)	0.815

Analysis in the presence of non-compliance

We next examined results based on schools' compliance with randomisation. Of the 66 intervention schools, 55 of them (83.3%) were found to have complied with the programme for RWI.[37] Further, we found that only 52 of the 65 control schools (78.5%) had not received portions of the treatment either by attending trainings or purchasing RWI materials. Table 17 provides the results from the CACE analysis for RWI using instrumental variables with details of the effect size calculations provided in Table C5 of Appendix C. The results for both primary and secondary outcomes are qualitatively similar to the primary and secondary analysis estimates reported in Tables 13 and 15.

Table 17: CACE analysis for RWI

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	**n (missing)**	**Mean (95% CI)**	**n (missing)**	**Mean (95% CI)**	**Total n (intervention; control)**	**Hedges' g (95% CI)**	**p-value**
NGRT – SAS	1,846 (705)	97.31 (96.64–97.97)	1,547 (816)	96.49 (95.77–97.21)	3,393 (1,846; 1,547)	0.06 (−0.01, 0.13)	0.525
KS1 writing passed – Year 3[38]	2,482 (39)	0.65 (0.63–0.67)	2,241 (26)	0.64 (0.62–0.66)	4,723 (2,482; 2,241)	−0.10 (−0.16, −0.05)	0.267

[37] As described previously, we used a generous definition of compliance whereby schools who received any aspect of the programme in the control group (including receiving materials or partial training) were considered to be in non-compliance with their random assignment, and treatment schools who did not attend training or receive any materials and, thus, did not implement RWI in their schools were considered non-compliant as well.

[38] The year group refers to the year pupils were in at the end of the evaluation, in summer 2018. The RWI pupils referred to as the Year 1 cohort were in Year 1 when the evaluation ended but began receiving the intervention while in Reception and completed

KS1 writing passed – Year 2	2,490 (61)	0.67 (0.65–0.69)	2,349 (14)	0.69 (0.67–0.71)	4,839 (2,490; 2,349)	−0.09 (−0.14, −0.03)	0.270
KS1 writing passed – combined	4,972 (100)	0.66 (0.65–0.67)	4,590 (40)	0.67 (0.65–0.68)	9,562 (4,972; 4,590)	−0.09 (−0.13, −0;05)	0.242
Phonics screening score – Year 2	2,484 (67)	33.50 (33.16–33.84)	2,363 (0)	33.27 (32.91–33.63)	4,847 (2,484; 2,363)	0.03 (−0.03, 0.08)	0.743
Phonics screening score – Year 1	2,014 (20)	35.14 (34.84–35.44)	1,685 (2)	33.68 (33.27–34.09)	3,699 (2,014; 1,685)	0.22 (0.16, 0.29)	0.001
Phonics screening score – combined	4,498 (87)	34.23 (34.00–34.47)	4,048 (2)	33.44 (33.17–33.71)	8,546 (4,498; 4,048)	0.12 (0.07, 0.16)	0.079
Passed phonics – Year 2	2,484 (67)	0.82 (0.80–0.83)	2,363 (0)	0.81 (0.80–0.83)	4,847 (2,484; 2,363)	0.03 (−0.03, 0.08)	0.782
Passed phonics – Year 1	2,014 (20)	0.89 (0.88–0.91)	1,685 (2)	0.84 (0.82–0.86)	3,699 (2,014; 1,685)	0.18 (0.11, 0.24)	0.004
Passed phonics – combined	4,498 (87)	0.85 (0.84–0.86)	4,048 (2)	0.83 (0.81–0.84)	8,546 (4,498; 4,048)	0.08 (0.04, 0.12)	0.139

As described above, the FS CACE analysis was unable to be conducted, as pupil compliance data could not be linked to outcome data from the NPD. We are therefore unable to explore the relationship between exposure to FS and reading attainment. Please see the *Fidelity* section and Appendix G for discussion of FS implementation.

outcome testing in Year 1. The RWI pupils referred to as the Year 2 cohort were in Year 2 when the evaluation ended but began receiving the intervention while in Year 1 and completed outcome testing in both Year 1 and Year 2. The RWI pupils referred to as the Year 3 cohort were in Year 3 when the evaluation ended but received the intervention and completed outcome testing while in Year 2.

Missing data analysis

Due to the extent of missing data, primary and secondary analyses for both RWI and FS were estimated on data completed using multiple imputation. Details on the analysis of missing data are provided in Appendix D. The results of the missing data analysis using imputed values are the same as our primary and secondary analysis for RWI.

Table 19: Imputed analysis for RWI

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	**n (missing)**	**Mean (95% CI)**	**n (missing)**	**Mean (95% CI)**	**Total n (intervention; control)**	**Hedges' g (95% CI)**	**p-value**
NGRT – SAS	2,534 (0)	96.14 (95.55–96.74)	2,380 (0)	96.23 (95.52–96.94)	4,914 (2,534; 2,380)	0.03 (−0.02, 0.09)	0.474
KS1 writing passed – Year 3[39]	2,437 (0)	0.63 (0.61–0.65)	2,351 (0)	0.66 (0.64–0.68)	4,788 (2,437; 2,351)	−0.06 (−0.12, −0.01)	0.238
KS1 writing passed – Year 2	2,534 (0)	0.67 (0.65–0.68)	2,380 (0)	0.70 (0.68–0.71)	4,914 (2,534; 2,380)	−0.04 (−0.10, 0.01)	0.275
KS1 writing passed – combined	4,971 (0)	0.65 (0.63–0.66)	4,731 (0)	0.68 (0.67–0.69)	9,702 (4,971; 4,731)	−0.04 (−0.08, −0.001)	0.226
Phonics screening score – Year 2	2,534 (0)	33.28 (32.93–33.63)	2,380 (0)	33.32 (32.97–33.67)	4,914 (2,534; 2,380)	0.01 (−0.04, 0.07)	0.780
Phonics screening score – Year 1	1,732 (0)	35.16 (34.83–35.49)	1,989 (0)	33.89 (33.52–34.26)	3,699 (1,710; 1,989)	0.18 (0.11, 0.24)	0.002
Phonics screening score – combined	4,266 (0)	34.04 (33.79–34.29)	4,369 (0)	33.58 (33.32–33.83)	8,613 (4,224; 4,369)	0.07 (0.03, 0.12)	0.092

Similarly, for FS, we find the results of the analysis qualitatively similar to the results from our primary and secondary analysis. Table 20 presents the full results of the analysis with imputed data.

[39] The year group refers to the year pupils were in at the end of the evaluation, in summer 2018. The RWI pupils referred to as the Year 1 cohort were in Year 1 when the evaluation ended, but began receiving the intervention while in Reception and completed outcome testing in Year 1. The RWI pupils referred to as the Year 2 cohort were in Year 2 when the evaluation ended but began receiving the intervention while in Year 1 and completed outcome testing in both Year 1 and Year 2. The RWI pupils referred to as the Year 3 cohort were in Year 3 when the evaluation ended but received the intervention and completed outcome testing while in Year 2.

Table 20: Imputed analysis for FS

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention; control)	Hedges' g (95% CI)	p-value
KS2 reading marks – Year 7[40]	684 (0)	19.87 (19.13–20.61)	615 (0)	22.31 (21.54–23.09)	1,299 (684;615)	−0.19 (−0.30, −0.08)	0.020
KS2 reading marks – Year 6	670 (0)	24.17 (23.33–25.01)	604 (0)	24.46 (23.63–25.29)	1,274 (670; 604)	−0.01 (−0.12, 0.10)	0.914
KS2 reading marks – combined	1,354 (0)	21.99 (21.41–22.58)	1,219 (0)	23.38 (22.81–23.95)	2,573 (1,354; 1,219)	−0.09 (-0.17, −0.02)	0.150
KS2 writing score − combined	1,354 (0)	94.18 (93.68–94.68)	1,219 (0)	94.21 (93.68–94.74)	2,573 (1,354; 1,219)	0.01 (−0.06, 0.09)	0.852

Subgroup analysis of ever FSM pupils

When limiting the sample to only those pupils who were ever FSM-eligible in the past 6 years, we found the magnitude of all effect sizes increased. Specifically, the impact of RWI on NGRT scores increased to +0.22 from +0.05 for all pupils but was not significant (p-value = 0.058). The interaction model found that the impact estimate on NGRT scores for FSM pupils was not significantly different from the estimate for non-FSM pupils. The effect size on KS1 writing passage increased from −0.06 and −0.04 to +0.08 and 0.00 for Year 3 pupils and Year 2 pupils, respectively, though these were not statistically significant (p-value = 0.319 and 0.981). Lastly, we found the estimated impact on phonics screening scores for Year 2 pupils increased from an effect size of 0.00 to an effect size of +0.06, but was not statistically significant (p-value = 0.374).

We were unable to test for subgroup differences for Year 1 pupils, as we did not have access to pupil-level covariates for this group. At the time of randomisation, pupils in the Year 1 cohort were in Nursery – a time when pupils are yet to be assigned UPNs. Accordingly, as UPNs were only collected for the trial at this one point in time, we were not able to match pupil examination scores to their characteristics for this cohort.

[40] The FS pupils referred to as the 'Year 7 cohort' were in Year 7 when the evaluation ended but received the intervention and completed the outcome testing while in Year 6. The FS pupils referred to as the 'Year 6 cohort' were in Year 6 when the intervention ended. They began receiving the intervention while in Year 5 and completed outcome testing in Year 6.

Table 21: Subgroup analysis for RWI – FSM only

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	**n (missing)**	**Mean (95% CI)**	**n (missing)**	**Mean (95% CI)**	**Total n (intervention; control)**	**Hedges' g (95% CI)**	**p-value**
NGRT – SAS	456 (279)	92.91 (91.57–94.25)	341 (159)	91.41 (89.83–93.00)	797 (456; 341)	0.22 (0.08, 0.37)	0.058
KS1 writing passed – Year 3[41]	810 (18)	0.54 (0.51–0.58)	606 (5)	0.52 (0.48–0.56)	1,416 (810; 606)	0.08 (−0.03, 0.19)	0.319
KS1 writing passed – Year 2	733 (2)	0.54 (0.50–0.57)	485 (15)	0.56 (0.52–0.61)	1,218 (733; 485)	0.00 (−0.11, 0.11)	0.981
KS1 writing passed – combined	1,543 (20)	0.54 (0.52–0.57)	1,091 (20)	0.54 (0.51–0.57)	2,634 (1,543; 1,091)	0.04 (−0.04, 0.12)	0.533
Phonics screening score – Year 2	724 (11)	31.14 (30.38–31.91)	482 (18)	31.16 (30.23–32.09)	1,206 (724; 842)	0.06 (−0.06, 0.17)	0.374
Passed phonics – Year 2	724 (11)	0.73 (0.70–0.77)	482 (18)	0.74 (0.71–0.78)	1,206 (724; 842)	0.02 (−0.09, 0.14)	0.814

Estimated effect sizes also increased in magnitude for FSM-eligible FS pupils. The estimated impacts on KS2 reading scores changed from −0.22 for all pupils to −0.26 for FSM pupils for the Year 7 cohort (those who received the FS intervention for only one year) and from −0.03 for all pupils to −0.13 for FSM pupils for the Year 6 cohort (those who received the FS intervention for two years). The overall combined effect size changed from −0.13 for all pupils to −0.21 for FSM pupils. However, only the effects on the Year 7 cohort and combined group were statistically significant for FSM pupils (p-value = 0.014 and 0.016, respectively). The impact estimate for KS2 writing scores combined for both Year 6 and Year 7 cohorts is qualitatively similar for all pupils and FSM pupils; the estimated effect size for all pupils was +0.02 and for FSM pupils only was +0.03 but was not statistically significant (p-value = 0.693). In other words, being in the FS treatment group resulted in greater learning losses for FSM pupils. For FS, pupils are taken out of their regular reading and writing classes to attend FS activities. One possible explanation for this negative effect could be that pulling pupils out of regular literacy classes is more detrimental to skills development for pupils who are already behind in their literacy development than having them attend regular literacy classes. Unfortunately, it is outside the scope of this evaluation to ascertain whether this was the case, so this explanation can only be taken as one potential interpretation of the negative impact on pupils assigned to FS programming. However, these results should be interpreted with caution as many schools assigned to implement FS programming did not actually provide the lessons to eligible pupils.

[41] The year group refers to the year pupils were in at the end of the evaluation, in summer 2018. The RWI pupils referred to as the Year 1 cohort were in Year 1 when the evaluation ended but began receiving the intervention while in Reception and completed outcome testing in Year 1. The RWI pupils referred to as the Year 2 cohort were in Year 2 when the evaluation ended but began receiving the intervention while in Year 1 and completed outcome testing in both Year 1 and Year 2. The RWI pupils referred to as the Year 3 cohort were in Year 3 when the evaluation ended but received the intervention and completed outcome testing while in Year 2.

Table 22: Subgroup analysis for FS – FSM only

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	**n (missing)**	**Mean (95% CI)**	**N (missing)**	**Mean (95% CI)**	**Total n (intervention; control)**	**Hedges' g (95% CI)**	**p-value**
KS2 reading marks – Year 7[42]	344 (32)	19.42 (18.43–20.40)	255 (34)	21.85 (20.72–22.99)	599 (344; 255)	−0.26 (−0.43, −0.10)	0.014
KS2 reading marks – Year 6	281 (54)	22.98 (21.72–24.24)	263 (39)	24.89 (23.66–26.12)	544 (281; 263)	−0.13 (−0.30, 0.03)	0.157
KS2 reading marks – combined	625 (86)	21.02 (20.23–21.81)	518 (73)	23.40 (22.55–24.24)	1,143 (625; 518)	−0.21 (−0.32, −0.09)	0.016
KS2 writing score – combined	696 (15)	92.96 (92.23–93.69)	578 (13)	92.96 (92.19–93.73)	1,274 (696; 578)	0.03 (−0.09, 0.14)	0.693

Subgroup analysis by gender

We also looked at estimated effects by gender to see if females were impacted differently than their male counterparts. Table 23 presents the RWI results for female pupils only. Estimated effect sizes for females were only slightly larger than those for all pupils, though none were statistically significant. The interacted model found that the impact estimate for female pupils was only statistically significantly different from the estimate for male pupils for NGRT scores and KS1 writing pass rates for Year 2 pupils. However, these differences were only marginally significant (p-value = 0.092 and 0.067, respectively).

[42] The FS pupils referred to as the 'Year 7 cohort' were in Year 7 when the evaluation ended but received the intervention and completed the outcome testing while in Year 6. The FS pupils referred to as the 'Year 6 cohort' were in Year 6 when the intervention ended. They began receiving the intervention while in Year 5 and completed outcome testing in Year 6.

Table 23: Subgroup analysis for RWI – females only

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	**n (missing)**	**Mean (95% CI)**	**N (missing)**	**Mean (95% CI)**	**Total n (intervention; control)**	**Hedges' g (95% CI)**	**p-value**
NGRT – SAS	825 (442)	99.74 (98.29–100/70)	871 (295)	98.50 (97.57–99.43)	1,696 (825; 871)	0.11 (0.01, 0.20)	0.185
KS1 writing passed – Year 3[43]	1,202 (16)	0.69 (0.66–0.72)	1147 (5)	0.72 (0.70–0.75)	2,349 (1,202; 1147)	−0.04 (−0.13, 0.04)	0.498
KS1 writing passed – Year 2	1,241 (26)	0.75 (0.73–0.77)	1149 (17)	0.76 (0.73–0.78)	2,390 (1,241; 1,149)	0.00 (−0.08, 0.08)	0.847
KS1 writing passed – combined	2,443 (42)	0.72 (0.70–0.74)	2296 (22)	0.74 (0.72–0.76)	4,739 (2,443; 2,296)	−0.02 (−0.08, 0.03)	0.603
Phonics screening score – Year 2	1,239 (28)	34.73 (34.34–35.12)	1161 (5)	34.29 (33.87–34.72)	2,400 (1,239; 1,161)	0.07 (−0.01, 0.15)	0.185
Passed phonics – Year 2	1,239 (28)	0.87 (0.85–0.89)	1161 (5)	0.86 (0.84–0.88)	2,400 (1,239; 1,161)	0.06 (−0.02, 0.14)	0.305

We further examined effects by gender for FS. Table 24 presents results for females only. Again, we found qualitatively similar results for all pupils and female pupils on all primary and secondary FS outcomes, though only the effect size on KS2 reading scores for Year 6 female pupils was found to be marginally significant. The interacted model found that none of the effects were statistically significantly different between females and males.

[43] The year group refers to the year pupils were in at the end of the evaluation, in summer 2018. The RWI pupils referred to as the Year 1 cohort were in Year 1 when the evaluation ended but began receiving the intervention while in Reception and completed outcome testing in Year 1. The RWI pupils referred to as the Year 2 cohort were in Year 2 when the evaluation ended but began receiving the intervention while in Year 1 and completed outcome testing in both Year 1 and Year 2. The RWI pupils referred to as the Year 3 cohort were in Year 3 when the evaluation ended but received the intervention and completed outcome testing while in Year 2.

Table 24: Subgroup analysis for FS – Females only

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	**n (missing)**	**Mean (95% CI)**	**n (missing)**	**Mean (95% CI)**	**Total n (intervention; control)**	**Hedges' g (95% CI)**	**p-value**
KS2 reading marks – Year 7[44]	254 (20)	20.15 (19.06–21.24)	242 (16)	22.79 (21.63–23.95)	496 (254; 242)	−0.23 (−0.40, −0.05)	0.052
KS2 reading marks – Year 6	242 (33)	24.61 (23.30–25.91)	221 (27)	25.78 (24.50–27.05)	463 (242; 221)	−0.06 (−0.25, 0.12)	0.516
KS2 reading marks – combined	496 (53)	22.32 (21.46–23.19)	463 (43)	24.22 (23.35–25.09)	959 (496; 463)	−0.12 (−0.25, 0.01)	0.161
KS2 writing score – combined	536 (13)	95.91 (95.18–96.64)	488 (18)	95.45 (94.65–96.25)	1,024 (536; 488)	0.09 (−0.03, 0.22)	0.259

Additional analyses and robustness checks

Additional analyses were conducted including the two control schools that failed eligibility requirements for the study after the randomisation had taken place. We excluded these schools from our primary analysis but include them here to test the robustness of our results. Table 25 reports the impact estimates for the primary and secondary outcomes for RWI inclusive of pupils from these two control schools. We found the results were qualitatively similar to our initial results from our primary and secondary analyses. Adding these two ineligible control schools back into our analytic sample did not substantially change the estimated effect sizes. Details of effect-size calculations are reported in Table C13 of Appendix C.

[44] The FS pupils referred to as the 'Year 7 cohort' were in Year 7 when the evaluation ended but received the intervention and completed the outcome testing while in Year 6. The FS pupils referred to as the 'Year 6 cohort' were in Year 6 when the intervention ended. They began receiving the intervention while in Year 5 and completed outcome testing in Year 6.

Table 25: Sensitivity analysis for RWI

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	**n (missing)**	**Mean (95% CI)**	**n (missing)**	**Mean (95% CI)**	**Total n (intervention; control)**	**Hedges g (95% CI)**	**p-value**
NGRT – SAS	1,653 (881)	97.04 (96.33–97.75)	1,759 (742)	96.78 (96.11–97.45)	3,412 (1,653; 1,759)	0.05 (−0.02, 0.12)	0.494
KS1 writing passed – Year 3[45]	2,403 (34)	0.63 (0.61–0.65)	2,443 (33)	0.66 (0.64–0.68)	4,846 (2,403; 2,443)	−0.04 (−0.10, 0.01)	0.360
KS1 writing passed – Year 2	2,492 (42)	0.67 (0.65–0.69)	2,460 (41)	0.70 (0.68–0.72)	4,956 (2,496; 2,460)	−0.04 (−0.10, 0.01)	0.272
KS1 writing passed – combined	4,899 (72)	0.65 (0.63–0.66)	4,903 (74)	0.68 (0.66–0.69)	9,802 (4,899; 4,903)	−0.04 (−0.08, −0.002)	0.295
Phonics screening score – Year 2	2,492 (42)	33.38 (33.04–33.73)	2,475 (26)	33.40 (33.06–33.74)	4,967 (2,492; 2,475)	0.02 (−0.04, 0.08)	0.662
Phonics screening score – Year 1	1,710 (22)	35.16 (34.83–35.49)	2,044 (15)	33.91 (33.54–34.27)	3,754 (1,710; 2,044)	0.18 (0.11, 0.24)	0.002
Phonics screening score – combined	4,202 (64)	34.11 (33.86–34.35)	4,519 (41)	33.63 (33.38–33.88)	8,721 (4,202; 2,519)	0.08 (0.04, 0.12)	0.071
Passed phonics – Year 2	2,492 (42)	0.82 (0.80–0.83)	2,475 (26)	0.82 (0.80–0.83)	4,967 (2,492; 2,475)	0.00 (−0.06, 0.06)	0.763
Passed phonics – Year 1	1,710 (22)	0.89 (0.88–0.91)	2,044 (15)	0.85 (0.84–0.87)	3,754 (1,710; 2,044)	0.15 (0.08, 0.21)	0.007
Passed phonics – combined	4,202 (64)	0.85 (0.84–0.86)	4,519 (41)	0.83 (0.82–0.84)	8,721 (4,202; 2,519)	0.05 (0.01, 0.10)	0.141

[45] The year group refers to the year pupils were in at the end of the evaluation, in summer 2018. The RWI pupils referred to as the Year 1 cohort were in Year 1 when the evaluation ended but began receiving the intervention while in Reception and completed outcome testing in Year 1. The RWI pupils referred to as the Year 2 cohort were in Year 2 when the evaluation ended but began receiving the intervention while in Year 1 and completed outcome testing in both Year 1 and Year 2. The RWI pupils referred to as the Year 3 cohort were in Year 3 when the evaluation ended but received the intervention and completed outcome testing while in Year 2.

Similarly, we conducted additional analyses including pupils from the two ineligible control schools for our FS analysis. Results of the additional analysis are presented in Table 26. Again, we found the point estimates for the effect sizes to be qualitatively similar to those found in our primary and secondary FS analyses. Adding these additional pupils into our analytic sample did not substantially affect the estimated impacts.

Table 26: Sensitivity analysis for FS

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	**n (missing)**	**Mean (95% CI)**	**n (missing)**	**Mean (95% CI)**	**Total n (intervention; control)**	**Hedges' g (95% CI)**	**p-value**
KS2 reading marks – Year 7[46]	637 (47)	20.34 (19.62–21.06)	585 (54)	23.08 (22.32–23.83)	1,222 (637; 585)	−0.25 (−0.36, −0.13)	0.006
KS2 reading marks – Year 6	579 (91)	24.74 (23.91–25.58)	573 (65)	25.68 (24.85–26.50)	1,152 (579; 573)	−0.04 (−0.16, 0.07)	0.593
KS2 reading marks – combined	1,216 (138)	22.44 (21.87–23.00)	1,158 (119)	24.36 (23.80–24.92)	2,374 (1,216; 1,158)	−0.15 (−0.23, −0.07)	0.046
KS2 writing score – combined	1,324 (30)	94.19 (93.68–94.70)	1,245 (32)	94.16 (93.64–94.67)	2,569 (1,324; 1,245)	0.02 (−0.06, 0.09)	0.832

[46] The FS pupils referred to as the 'Year 7 cohort' were in Year 7 when the evaluation ended but received the intervention and completed the outcome testing while in Year 6. The FS pupils referred to as the 'Year 6 cohort' were in Year 6 when the intervention ended. They began receiving the intervention while in Year 5 and completed outcome testing in Year 6.

Implementation and process evaluation

Implementation

In this section, we present findings from qualitative data collected by AIR on programme implementation in two treatment schools, business-as-usual in one control school, and 113 results from the fidelity to teaching survey. We conducted three key informant interviews with RMT staff, four key informant interviews with school leadership/reading leaders, and two focus groups with teachers and teaching assistants involved in the delivery of the programme (total of 6 key informants and 12 focus group participants). In addition, we conducted approximately a dozen rapid observations of RWI and FS lessons in two programme schools. We focus our analysis on perceptions of necessary conditions for the successful implementation of the programme, as well as barriers to delivery and implementation challenges.

Necessary conditions for successful implementation

Across our qualitative interviews, focus groups and observations, three elements emerged as necessary conditions for successful implementation of the programme: getting buy-in from school leadership (particularly headteachers), establishing good relationships and communication between RWI trainers and school staff, and integrating practice sessions into school routines.

Fostering buy-in with school leadership

According to RMT staff, headteachers and reading leaders, fostering buy-in with school leadership, particularly headteachers, is an important element to successful implementation of RWI and FS programmes. Implementing RWI and FS requires a significant amount of logistical support from headteachers, both in terms of managing teaching staff needs and devoting time and resources for the continuous professional development of programme instructors. For example, since there are usually a larger number of assessment-based RWI and FS groups than teachers, headteachers must use trained teaching assistants to lead some of the groups. This factor can create a logistical challenge because teaching assistants must leave their usual classrooms to teach RWI groups. In addition, some teaching assistants reported being inexperienced in leading classes, making it all the more important for headteachers to provide continuous professional development for programme instructors. For example, a RMT staff member stated,

> *"The more we have access to the [school's] leadership team and the more we can help schools implement their own internal weekly [continuous professional development] – where they practise and coach, the better they will be."*

A RMT trainer echoed this view, adding that headteachers must be engaged with all aspects of the programme in order to provide adequate support for teachers. She explained,

> *"It's best (…) when the headteacher is with me on the journey the whole way, that means they prioritise coming with me to development days, they come and see the lesson coaching from the development day, they're there for the whole of the training. If the headteacher doesn't get it then they're not going to give the staff the time that's required to become expert teachers."*

According to the fidelity to teaching survey, only 33% of headteachers surveyed reported working alongside trainers on development days, suggesting that many headteachers may not be as involved in the day-to-day implementation of the programme.

Data from headteachers and reading leaders also supported this finding. A headteacher explained that it is crucial to provide the time and support for the reading leader and programme instructors to prepare for RWI sessions alongside their other responsibilities. He explained,

> *"[the] head teacher has to commit to it and that means providing time for your reading leaders and providing time for the practice sessions and (…) for the teachers and the teaching assistants to attend those and then making it a non-negotiable."*

To illustrate, he described how he provided cover for reading leaders to do coaching and built time for RWI prep and practice sessions into the teaching assistants' working hours. Further, a reading leader from a different school noted

that having the support of her headteacher was essential for her to have the time to analyse RWI data and adequately group the children, alongside her other teaching responsibilities. According to the fidelity to teaching survey, 53% of reading leaders reported being released from the timetable to coach and support RWI activities. In addition, most reading leaders reported that they were only given two hours or less per week to perform duties for RWI lessons and other RWI-related tasks. These findings suggest that headteachers in many programme schools were not as involved in development days as expected, and that reading leaders in almost half of programme schools were not released from the timetable for enough time to support RWI activities. Fostering buy-in with school administrators may be an important factor in ensuring that reading leaders can adequately support teachers and carry out key components of the programme, such as grouping and assessing children.

Maintaining frequent communication between trainers and school leadership

Our qualitative data suggests that maintaining frequent communication between trainers and school leadership outside of development days was also an important element of successful programme implementation. The professional development aspect of the RWI programme included periodic development days where trainers visited programme schools, analysed data alongside headteachers and reading leaders, and offered on-the-spot coaching to programme instructors. Our data indicates that frequent communication between trainers and reading leaders outside of development days facilitated programme implementation. RMT trainers explained that frequent contact with reading leaders outside of development days was helpful in providing troubleshooting and real-time feedback on issues such as grouping. For instance, a trainer stated,

> *"I create such strong relationships with my reading leaders that they will email me or phone me constantly, and that is the way forward because if they see you as part of the team then you know what's going on. [Because] (…) if you don't have that contact then you don't know if they're teaching with fidelity and that's when you get a surprise, that's when the results might not be as good."*

Another trainer explained that having frequent email and phone contact with reading leaders made it easier to address implementation and fidelity issues as they emerged, rather than waiting to address them during development days. This finding was echoed by the headteachers and reading leaders interviewed for this study, who stated that developing a strong relationship with trainers and feeling comfortable reaching out to them was an important element in addressing grouping issues and adapting to the needs of specific children. In addition, a headteacher stated that having the same trainer helped develop this sort of relationship over time.

School support for regular practice sessions

Lastly, we identified supporting regular practice sessions in schools as key to ensuring high-quality programme delivery. Beyond trainings and development days, schools are also encouraged to support continuous professional development by holding weekly RWI practice sessions with teachers and teaching assistants led by the reading leader. These practice sessions depended on the headteacher's commitment and management of the time and workload of teachers. As a trainer explained,

> *"it's all down to getting the headteacher on board so that there are practice sessions, because if you don't have practice sessions the teaching assistants are left alone to give it their best shot, and if somebody isn't supporting them then you won't bring up the teacher's proficiency level."*

One of the programme schools visited by AIR gave time for teachers and teaching assistants to conduct RWI practice sessions once every two weeks. Teachers and teaching assistants listed it as an important part of becoming proficient in the programme. A teacher noted,

> *"Because then, you start to put it into practice (…) it starts to click, doesn't it, (…) and gradually, the rough edges, uh, get knocked off a bit."*

As teachers become proficient in the programme, some schools may deem practice sessions unnecessary. In one of the schools visited for this study, practice sessions were discontinued in the second year of the trial because school leadership and teaching staff believed they were proficient enough with the programme. Only 40% of schools in the teaching to fidelity survey sample reported having weekly practice sessions by reading leaders. The lack of regular

practice sessions may be related to teachers and teaching assistants successfully becoming proficient in the programme, or to the issue of reading leaders not having enough time release to conduct programme activities.

Barriers to delivery and implementation challenges

According to our interview and focus group respondents, some elements of the programme were challenging for schools to implement, particularly for FS, including a lack of awareness on the part of schools about the details of the FS programme, inadequate FS materials, lack of structure in the FS programme, and lack of FS-specific training for programme instructors. According to RMT staff, this may be due to gaps in communication between QUB and schools during the recruitment process for the trial. In regard to the RWI programme, some headteachers and reading leaders have found it challenging to meet the staffing and space-related needs of the programme. Additionally, some interview and focus group respondents noted challenges that may affect teachers' ability to implement the RWI programme as intended. For example, some teaching assistants didn't feel adequately supported to deliver the RWI programme, and a few teachers and teaching assistants noted that they struggled to maintain the pace of RWI lessons and effectively integrate writing in the programme scheme.

Concerns with training and resources for FS

Fidelity to teaching survey data and qualitative data from six interviews in two treatment schools indicate that the FS programme was not being delivered as intended in some programme schools. The fidelity to teaching survey found that FS was only delivered in 44% of programme schools in 2017/18, while most of the remaining schools reverted to business-as-usual, and a small proportion delivered RWI to FS-eligible pupils in Year 5 and 6 instead. In one of the schools visited by AIR, FS had been discontinued in the second year of the trial. According to the headteacher, this was because all the FS pupils had passed the phonics assessment and graduated out of the programme. At the same time, teachers and teaching assistants at the same school noted that FS suffered from numerous challenges. For instance, teachers and teaching assistants were not trained on FS specifically and felt that the RWI training they received did not adequately prepare them for FS implementation.[47] As one teacher explained,

> *"Read Write Inc. is very comprehensive, and anybody theoretically can do it, if they can read themselves, they can do it. But Fresh Start, […] if you're not trained to do that anyway, it's quite difficult for a TA to pick that up and just go running with it, and know what to do."*

Further, teachers and teaching assistants from both schools in our sample reported that the FS handbook[48] had gaps and was not as comprehensive as the RWI handbook. The reading leader from the other selected school (which continued to implement FS), thought that the recently revised FS handbook was more helpful, but noted that it was a nuisance to have to acquire all the new FS materials again during the second year of the trial. Data from interviews with RMT staff indicate that many programme schools experienced challenges implementing FS, and suggest that this may be due to a lack of clarity on the specifics of the FS programme during the initial recruitment of schools for the EEF trial. RMT staff believed that the recruitment process did not adequately communicate the expectation to implement FS and the full details of FS implementation to schools who were signing up for the trial. However, the four school-level qualitative interviews and two focus groups we conducted in two treatment schools did not indicate recruitment was an issue.

Staffing and space needs

Implementing RWI creates unique operational challenges for school leadership. Since the programme entails grouping children into several levels based on assessments, it often requires teaching assistants to be trained and support to

[47] Data from interviews with RMT staff indicate that there is a separate FS training, but it is usually delivered only to high-need schools. Most programme schools only get the RWI phonics training.
[48] Interviews with RMT trainers and reading leaders indicate FS materials had been recently revised. However, the school that stopped FS implementation had not acquired the revised handbook.

teach groups of children. In addition, the work of assessing and grouping children every half term demands time from reading leaders. One headteacher summarised the issue as follows,

> *"If you're going to run [the programme] effectively, it's very child-centred in that you're carrying out the assessment and therefore the children need to be assigned the correct reading group. You can end up having more groups than you have adults, which is a problem."*

Although RMT trainers work with headteachers and reading leaders to resolve these issues, they still seem to be a challenge for the two programme schools we visited.

Both programme schools visited by AIR met the staffing need by having teaching assistants teaching RWI and FS groups. According to trainers, this is also the case in most programme schools. These staffing needs can pose logistical challenges, especially for small schools. For instance, teaching assistants must leave their own classes to teach RWI lessons, and the demand for teaching assistants can change as children get reassessed into different groups throughout the year. As one reading leader put it,

> *"We've got TAs that are in all different year groups that come to deliver Reception or Year 1 and 2 Read Write Inc. (…) if they're with Read Write Inc. it means they're not in class (…) And it might be that next half term, they're not needed and they can go back, and the next half term they might need to come back again. So as the groups change, the adults change, I think that's the biggest challenge really, the adults."*

A RMT trainer noted that staffing was more of an issue in small schools with fewer staff available to teach the programme. She explained,

> *"If it's a small school sometimes we have to really think out of the box cause they don't have enough staff. So they'll recognise that maybe they've got 7 groups, you'd ideally have 7, but you only have 5 teachers (…) I find it much easier in a big school where I've got a lot more staff to play with."*

In addition, finding space to conduct multiple RWI lessons simultaneously can also be an issue. In the schools visited by AIR, some RWI lessons had to be conducted in makeshift locations, such as hallways. In one of the schools, these makeshift locations were reported to be more subject to distractions, such as noise and passers-by. Further, teachers and teaching assistants noted that some spaces used by RWI did not have access to internet and projectors, affecting their ability to use online programme resources.

Teaching assistants require support from the reading leader

Further, respondents noted additional challenges that may affect their ability to adequately deliver RWI lessons to pupils. For instance, many teaching assistants interviewed in focus groups in the two treatment schools visited by AIR noted that they did not feel adequately supported in delivering the RWI programme. RWI implementation required a shift in expectations around teaching assistant responsibilities from a support role to a more active teaching role. As one headteacher explained,

> *"[There has] been a shift in terms of thinking about what the expectations are around teaching assistants and what levels of accountability they have then for the progress of their pupils (…) that was a challenge for us to develop the confidence for the teaching assistants to lead (…) and for them to also accept that there's a level of accountability that comes with it."*

In focus group discussions, many teaching assistants noted that they lacked teaching experience prior to teaching RWI and FS, both in terms of pedagogy and classroom management. This, coupled with the process of becoming proficient in programme content and delivery, created a steep learning curve for many teaching assistants involved with RWI. The lack of support for teachers and teaching assistants to overcome this learning curve was noted as a challenge to implementing RWI.

Although most teaching assistants who participated in focus groups in the two schools visited by AIR agreed that the RWI training and continuous professional development had been effective in teaching them RWI programme content, many described needing more support in classroom and behaviour management. As one teaching assistant explained,

> *"We never really [were] told how (…) to manage the children, class management and all that. (…) We were just chucked straight into the deep end, you know, you've never taught a class before, here you go, here is 24 kids, go and teach them for an hour. (…) All that sort of stuff isn't put into the training – we just get taught how to deliver Read Write Inc. We don't get taught how to manage children and how to deal with different situations."*

This demand was not necessarily directed at the RWI programme itself – as one teacher mentioned, school leadership could also step in to fill that need. In addition, some teaching assistants felt that they were not given adequate time to prepare RWI lessons along with their other responsibilities. One teaching assistant described,

> *"We're leaving our Year 4 or Year 5 class, to do [RWI], and then we're kind of back to our classes, and (…) everything we had to do before we left still needs to be done, and we've lost an hour of our day, plus, extra time in between."*

RWI programme instructors struggle with pacing and teaching writing

In addition, many teachers from both treatment schools in our qualitative sample noted challenges in keeping up with the expected pace of RWI lessons, which affected the time they spent on writing. Depending on the level, RWI lessons consist of three- or five-day programmes, where pupils learn new sounds, read a story that contains those sounds, and teach a writing task at the end to practice compositional writing. For the writing task, pupils have to put together multiple skills, including decoding, vocabulary, composition and writing. Some teachers and teaching assistants felt that the pace of the RWI programme was too fast for their pupils, noting that they struggled to spend enough time on writing. Further, some teachers felt that they needed to teach elements of writing that were not accounted for in the RWI scheme, such as the mechanics of writing or genre. As one teaching assistant explained,

> *"[In the blue group, at] the top end of the Read Write Inc. scheme, the children actually cannot write on a line. So you're actually having to teach although it's not mentioned in the handbook, if you follow the handbook to the T. Um, the writing part takes up a chunk of the time, and it's just part of the whole programme. And the blue group, and the orange as well, is a five-day programme. Sometimes we have to take – I let it overlap into six days."*

In response, teachers and teaching assistants reported adapting the programme length and/or adding activities that bring more context and purpose to reading and writing (props, visual aids, and so on).

Fidelity

Table 27 presents teachers' reported attendance and participation during the RMT trainings for both RWI and FS. There was a higher reported frequency of the reading leader working alongside the trainer than the headteachers working alongside the trainer in both years, with headteachers significantly reducing their role working alongside the trainer in 2017/18 (55% to 33%). Although overall staff attendance was high (about 90% said they attended), the attendance of the headteacher throughout the training occurred at only 68% of the schools across both school years. The questions included in the fidelity measures reflect what RMT defines as proper implementation of RWI and FS.

Table 27: Training: attendance and participation

Variable	Overall		SY 2016/17			SY 2017/18			Difference in means (7) − (4)	p-value of difference
	N[49] (1)	MEAN (2)	N (3)	MEAN (4)	SD (5)	N (6)	MEAN (7)	SD (8)		
All staff attend the two-day phonics training attendance and participations	113	0.88	58	0.914	0.283	55	0.855	0.356	−0.059	0.33
Headteacher attends throughout the two-day phonics training attendance and participations	113	0.68	58	0.672	0.473	55	0.691	0.466	0.019	0.83
Headteacher works alongside trainer on development days	113	0.44	58	0.552	0.502	55	0.327	0.474	−0.225	0.02
Reading leader works alongside trainer on development days	113	0.96	58	0.966	0.184	55	0.945	0.229	−0.021	0.61

Table 28 highlights the central role played by the reading leader in the trainings and implementation of the programme. Although 30-minute coaching sessions for staff conducted by reading leader occurred in fewer than half the schools each school year (about 40–45%), the reading leader was actively involved in most schools in terms of other activities. This was despite the reading leader being less likely to be released from their timetable to coach and support pupils (decreased from 74% to 53%). Lastly, almost half of the schools in 2017/18 reported a change to either the headteacher or reading leader since the previous school year. However, reading leader involvement remained high, suggesting these changes had limited impacts on the fidelity of programme implementation at the school level.

Table 28: Reading leader coaching/involvement during program

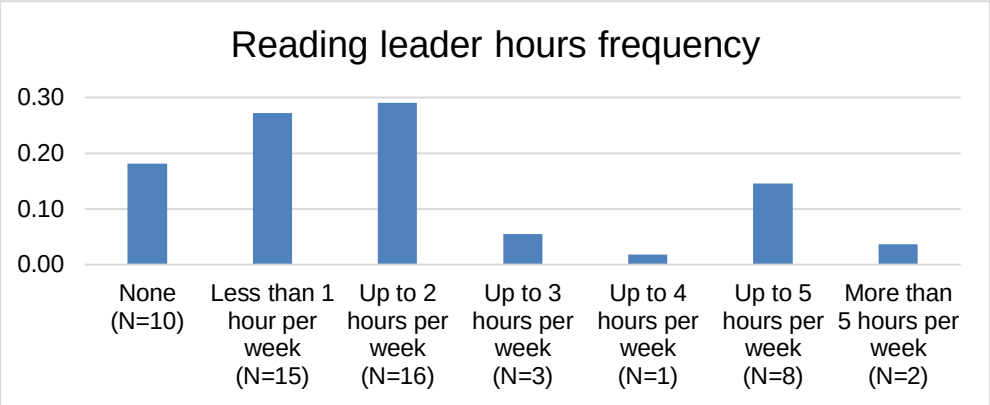
Variable	Overall		SY 2016/17			SY 2017/18			Difference in means (7) − (4)	p-value of difference
	N (1)	MEAN (2)	N (3)	MEAN (4)	SD (5)	N (6)	MEAN (7)	SD (8)		
Reading leader coaches staff in weekly 30-minute sessions	113	0.42	58	0.448	0.502	55	0.400	0.494	−0.048	0.61
Reading leader released from timetable to coach and support	113	0.64	58	0.741	0.442	55	0.527	0.504	−0.241	0.02
Reading leader assesses phonics and FS children using RWI assessments every half-term	113	0.92	58	0.879	0.329	55	0.964	0.189	0.085	0.17
Reading leader places children in close homogeneous groups	113	0.96	58	0.948	0.223	55	0.964	0.189	0.016	0.69
Reading leader organises daily 10-minute one-to-one tutoring for all children not making expected progress	113	0.81	58	0.793	0.409	55	0.836	0.373	0.043	0.56

[49] N in Tables 27–31 refers to the number of schools, as one teacher from each treatment school was asked to complete the fidelity survey each school year.

Any changes to reading leader or headteacher	55	0.46	–	–	–	55	0.455	0.503	–	–

In the 2017/18 school year, schools were asked about the number of hours reading leaders were given to perform their duties for both RWI lessons and other tasks. Figure 6 shows that reading leaders from the majority (75%) of schools were given 2 hours or less of time to perform all of their duties per week.

Figure 6: Reading leader hours



Tables 29 and 30 provide the summary statistics for the implementation of the RWI programme and the FS programme, respectively. Although the level of detail of timing that the programme is implemented is quite granular, the results illustrate that the programme was mostly being implemented as intended according to these teacher self-report survey questions. The proportion of schools reporting teaching RWI for at least an hour decreased significantly from last year (from 74% to 53%). The minimal impacts of RWI on pupils' reading outcomes may be attributed that the low levels of implementation fidelity, especially on outcomes measured in the second year of the programme (when fidelity decreased).

Table 29: Reception and Year 1 children receiving programme (RWI)

Variable	Overall		SY 2016/17			SY 2017/18			Difference in means (7) − (4)	p-value of difference
	N (1)	MEAN (2)	N (3)	MEAN (4)	SD (5)	N (6)	MEAN (7)	SD (8)		
Reception children are taught phonics daily for a minimum of 20 minutes in Term one	113	0.96	58	0.931	0.256	55	1.00	0.00	0.069	0.04
Reception children are taught in homogeneous groups after the first 4 weeks of teaching	113	0.88	58	0.845	0.365	55	0.927	0.262	0.082	0.17
Reception children are taught for a minimum of 30 minutes in Term 2 and 40 minutes in Term 3	58	0.95	58	0.948	0.223	–	–	–	–	–
Phonics lessons are taught daily for one hour in Year 1 upwards	113	0.64	58	0.741	0.442	55	0.527	0.504	−0.214	0.02

Table 30: FS children receiving programme

Variable	Overall		SY 2016/17			SY 2017/18			Difference in means (7) − (4)	p-value of difference
	N (1)	MEAN (2)	N (3)	MEAN (4)	SD (5)	N (6)	MEAN (7)	SD (8)		
FS is taught in Years 5 and 6	113	0.51	58	0.586	0.497	55	0.436	0.501	−0.150	0.11
FS lessons are taught daily for one hour	113	0.28	58	0.345	0.479	55	0.218	0.417	−0.127	0.14

Usual practice

For this section, we draw on information collected by QUB through MoUs and qualitative data AIR collected from control and programme schools.

Most control schools included in the MoU data reported using Letters and Sounds to teach phonics, either on its own or in combination with other programmes. One-third of control schools reported using a combination of Letters and Sounds and Jolly Phonics. Additionally, one out of five control schools reported using RWI in some way, either on its own, or combined with Letters and Sounds. Other existing phonics approaches reported less often by control schools included Batty Phonics, Phonics Play, and Floppy's Phonics.

Virtually all control schools reported using a combination of different programmes and approaches to teach reading. The most commonly cited approach to teach reading was Oxford Reading Tree (over half of schools), followed by RWI/FS, Rigby Star, and Collins Big Cat. Other approaches to teaching reading that were cited less often by schools included Bug Club, Code X, Project X Code, and Reading Recovery. In addition, some schools mentioned using different types of book banding approaches and a few schools reported not having a specific model to teach reading.

Control schools reported a wide range of interventions and schemes for low attaining readers in Years 5 and 6, citing over 27 different interventions or teaching strategies. Schools tended to report using a combination of different interventions and strategies rather than using one intervention on its own to teach low attaining readers in Years 5 and 6. No one intervention dominated across control schools, but some of the most commonly cited included some commercially available interventions, such as Project X Code, FS, Rapid Reading, Lexia, and Reading Recovery. The most commonly cited teaching strategies for low attaining readers of this age included one-to-one reading sessions with teaching assistants or teachers and guided reading sessions. In addition, a few schools noted that they did not use a particular intervention to teach low attaining readers in Years 5 and 6, but teachers developed individualised interventions based on the needs of particular children.

The data AIR collected from the selected control school seems to be consistent with the trends that emerged from the MoU data. According to administrators, the school used a variety of approaches to teaching reading. In Nursery, the school begins to teach reading through games and phonics, with a strong emphasis on reading a wide range of texts to children. In Reception, all children start Phase 2 phonics from the Letters and Sounds programme. From Reception to Year 2, children progress through the Letters and Sounds scheme at their own pace in ability-based groups. However, teachers are granted autonomy to use other phonics schemes beyond Letters and Sounds, although Letters and Sounds is recommended. For example, the school also uses the book banding approach and some RWI spelling resources. Children who are not progressing through the phonics scheme are identified by the teacher and given extra phonics intervention at the teacher's discretion, with the help of teaching assistants. After Year 1, the focus shifts from phonics to reading comprehension. After Year 2, children participate in guided reading groups. If a child is not reading by Year 3 or 4, the school obtains additional professional support through an education psychologist or a specialist in learning difficulties. Further, struggling readers may receive one-to-one reading sessions with adults. There were no interventions or schemes equivalent to FS being implemented by the school.

The two treatment schools visited by AIR used the Letters and Sounds scheme, as well as an approach inspired by Reading Recovery and guided reading groups before the EEF trial. The headteacher from one of the schools noted that though they taught phonics previously, the approach to teaching wasn't as structured and consistent as it needed to be in the early years. One school was also using Letters and Sounds, with support from Jolly Phonics resources before RWI. The headteacher from the other school explained that while Jolly Phonics was a step in the right direction from

the previous approach (introducing a letter a week to Reception pupils), children could still slip through the cracks, and it took longer to pick up on children that were not doing as well. In this headteacher's view, RWI made it easier to identify children who were struggling with reading. AIR has no knowledge of other relevant phonics interventions being carried out in these two treatment schools during the time of the programme.

Cost

Using the assumptions listed earlier (each teacher uses the programme with three years of pupils), the financial costs of the RWI intervention were £17,460 per school for three years (£5,820 per school per year) and £171 per pupil per year for schools with a High Ofsted score, or £20,460 per school for three years (£6,820 per school per year) and £201 per pupil per year for schools with a Low Ofsted rating (Table 31). On average, the cost of the RWI intervention was £18,960 per school for three years (£6,320 per school per year) and £186 per pupil per year. We also assess the relative cost-effectiveness of RWI by dividing the per school and per pupil costs by the standardised gain in NGRT scores. This analysis found that, for schools with High Ofsted scores, it would cost £116,400 per school per year or £3,423 per pupil per year for one month's progress, and for those with Low Ofsted scores it would cost £136,400 per school per year or £4,011 per pupil per year for one month's progress as they required 5 additional development days. On average, the relative cost-effectiveness of RWI for one month's progress is £126,400 per school per year, or £3,718 per pupil per year.

Table 31: Cost of delivering RWI to one school

Item	Type of cost	Cost per school	Total cost per school over 3 years	Total cost per pupil per year over 3 years
Headteacher and reading leader training days: 2 sessions in 1 year	Start-up cost for each headteacher and reading leader	£520	£520	
Regional training for headteacher and reading leader: 2 places over 1 year	Start-up cost for each headteacher and reading leader	£540	£540	
In-school staff training: 2 sessions in one year	Start-up cost for each school	£2,200	£2,200	
Development days: 6 over 2 years for High Ofsted; 11 over 2 years for Low Ofsted	Running costs for the programme, incurred once per school	£3,600 for High Ofsted £6,600 for Low Ofsted £5,100 on average	£3,600 for High Ofsted £6,600 for Low Ofsted £5,100 on average	
Annual supply costs to cover teachers attending training and teaching RWI	Running costs of the programmed, incurred once per school	£4,200	£4,200	
Annual costs for each teacher	Running cost per teacher	£0	£0	
Initial RWI materials cost	Start-up cost for each school	£4,000	£4,000	
Annual RWI materials cost	Running cost per school	£800	£800 × 3 = £2,400	

Total			£17,460 for High Ofsted £20,460 for Low Ofsted £18,960 on average	(£17,460/3/34) = £171 for High Ofsted (£20,460/3/34) = £201 for Low Ofsted (£18,960/3/34) = £186 on average

In this evaluation, 131 schools and 4,454 Year 2 pupils were recruited into the study – about 34 pupils per school. These numbers were used to calculate the training costs per school and final costs per pupil. The intervention schools incurred initial costs related to training conducted by RMT in order to implement the RWI programming. These training sessions included the following:

- Regional phonics trainings for the headteacher and reading leader;
- In-school phonics training for all staff (up to 45 staff);
- Development days based on a school's Ofsted score;
- Regional training for headteacher and reading leader on reading leader training Day 1 and Day 2.

Once trained, schools incurred no staff costs for using the RWI programme in subsequent years. We assume that, once trained, teachers would use RWI for the following three years – this follows the standard approach for calculating cost in EEF trials and is not an assumption specific to this programme. The only non-financial costs included teacher time to prepare for and deliver RWI sessions.

Schools incurred one-time costs associated with the purchase of RWI materials from Oxford Publishing, including flash cards, posters and workbooks. Additionally, schools must repurchase workbooks each year for each new cohort of pupils to participate in RWI sessions.

The costs incurred by schools participating in this evaluation were slightly lower than the costs presented in Table 31 (those that will be incurred to schools implementing RWI in the future). Schools in our study had costs supplemented by EEF such that schools received a 20–50% discount on all training costs through RMT. Additionally, RMT offers discounted training bundles schools can purchase which provide a slight discount to schools over purchasing the trainings a la carte.

Conclusion

Table 32: Key conclusions

Key conclusions
Children in Read Write Inc. Phonics (RWI) schools made the equivalent of one additional month's progress in reading, on average, compared to children in other schools. This result has a low to moderate security rating.
Children in Fresh Start (FS) schools made the equivalent of two months' less progress in reading, on average, compared to children in other schools. This result has a moderate security rating.
Children eligible for Free School Meals (FSM) in the RWI intervention group made the equivalent of three months' additional progress in reading, on average, compared to children in other schools, while children eligible for FSM in the FS intervention group made the equivalent of three months' less progress, on average. Both of these findings must be interpreted with caution, as they are based on small subgroups of pupils eligible for FSM.
RWI had a positive impact on pupils' phonics outcomes at the end of Year 1, equivalent to one month's additional progress. This was the average impact for pupils who received the programme for one year and pupils who received the progamme for two years before taking the Phonics Screening Check. Neither children in the RWI or FS intervention group made more progress in writing than pupils in other schools.
FS was not implemented as intended in a significant proportion of intervention schools. 23 out of 66 intervention schools (35%) did not deliver FS at all and 19 schools (29%) delivered FS to some but not all eligible pupils. The low levels of engagement with FS will have influenced the impact results for this programme, so the results should be interpreted with caution. There was greater engagement with RWI, although 10 schools (15%) did not implement the programme.

Impact evaluation and IPE integration

Interpretation

The results of this evaluation showed that RWI had a modest positive impact of one month's additional progress on pupils' reading and 0 months' additional progress on writing knowledge attainment, on average. The programme only led to small increases in pupils' phonics outcomes for those in the Year 1 cohort (pupils who received the programme for two years before taking the test) and slight, though non-significant, decreases on writing outcomes regardless of programme exposure. For RWI, the longer dosage of programme exposure for Year 1 pupils produced much larger results than for the Year 2 pupils who only received the programme for one year prior to taking the Phonics Screener at the end of Year 1 (+0.18 ES compared to +0.01 ES).[50] For the FS programme, there was evidence that being allocated to the FS treatment group had a negative impact on the Year 7 cohort's remedial reading (pupils who received the programme for only one year, while they were in Year 6) and no impact on pupils' writing knowledge regardless of dosage.[51] In fact, based on our primary analysis, pupils who were randomly selected to receive the FS programme performed worse on national reading and writing assessments, on average, than pupils who received business-as-usual. While it is important to recall that 23 schools (35%) in the treatment group did not deliver FS programming to eligible pupils while 19 schools (29%) in the treatment group delivered FS to some but not all eligible pupils, the implementation fidelity issues alone cannot account for the negative results we find. However, while we can offer suggestions as to what may have led to the negative impact (e.g., removing pupils from regular literacy classes, causing already struggling pupils to fall further behind), this trial does not provide us with firm evidence to explain what the difference was between intervention schools and control schools which accounts for this result. The implementation and process evaluation constituting survey data from 58 treatment schools and interview data from two treatment schools revealed that, while RWI programming was implemented with relatively high fidelity, FS programming was not well implemented due to resource constraints, lack of specific training on FS programming, lack of a clear structure in the FS programming, and a lack of materials.

[50] The year group refers to the year the pupils were in at the end of the evaluation.
[51] The year group refers to the year the pupils were in at the end of the evaluation.

In terms of our theory of change, we hypothesised that a key step along the causal linkages between the trainings and assessments and reading and writing outcomes is that phonics lessons are implemented with fidelity. Although the fidelity to teaching surveys done by the programme staff themselves did point to high levels of implementation fidelity of the RWI programme, there was clearly much less implementation fidelity of the FS programme. Fostering a sense of buy-in from school authorities along with continuous support for the regular practice sessions was considered critical for the implementation of the programme with fidelity. Data from the four key informant interviews and two focus group discussions also highlighted that a lack of physical space and staffing support needed to implement the lessons were also felt by the schools, highlighting the heavy resource requirements of the programme, and the difficulty in integrating the time-intensive programme with regular programmes. Due to the low levels of implementation fidelity for FS, we caution that the results of our trial may not accurately reflect the impact the programme could have had on pupils' remedial reading outcomes if delivered with fidelity. Even so, the objective of the ITT analysis is such that it ascertains the impact of an intervention based on the assignment to treatment conditions rather than with respect to compliance with random assignment. Therefore, results of our ITT analysis still provide useful insights, especially in tandem with contextual factors regarding the fidelity of implementation of the FS programme.

Overall, RWI shows some promise in supporting early phonics and decoding skills of the reading acquisition process but needs consistent support for successful implementation. FS, on the other hand, requires additional evaluation to determine its effectiveness.

Limitations and lessons learned

There are a number of limitations to this study that should be considered when interpreting results.

First, the low fidelity and inconsistency of implementation for FS programming in intervention schools limits our ability to fully assess the impact of the FS programme. While the CACE analysis would have aimed to circumvent this problem, we were unable to identify eligible FS pupils who actually received the programming due to data limitations, though we are aware that only a small proportion of FS-eligible pupils actually received the intervention. Therefore, it has not been possible to produce a reliable CACE analysis for FS. Furthermore, our sample size for FS was relatively small, such that we were not as powered to detect meaningful effect sizes. Even so, our analyses detected some significant effects. A larger trial focusing only on administration of FS programming would improve the ability to assess the intervention and elicit any valid claims about the programme's effectiveness, although given the complexities and low levels of implementation fidelity in this evaluation, it is questionable whether it would be possible to run at a larger scale.

Second, the distribution of the NGRT SAS showed clear floor effects, rendering our analysis incapable of accurately measuring pupils' true average reading scores, especially for those at the bottom of the distribution. This result suggests the assessment used to measure our primary outcome for the RWI analysis may have been too difficult for many pupils. Therefore, the secondary outcomes may be more beneficial in assessing the effectiveness of the RWI intervention.

A third limitation of this study is its reliance on self-reported data. While we triangulated reports across different respondents, we were not able to fully analyse how respondents' reported practices compared to their actual practices. The timeline of this study did not allow for the collection of systematic observation data, which would have enabled us to triangulate participants' views and practices.

Fourth, the implementation and process evaluation suffered from non-representativeness. Due to the limited scope of this light touch process snapshot, we were only able to visit two programme schools. We purposively selected schools according to performance and geographic region. While this approach allowed us to gain knowledge on the implementation of the RWI and FS programmes in different contexts, it was not designed to generate a representative sample. We also have very limited control to know what was happening in these schools. Therefore, the findings of the qualitative process snapshot are not generalisable to all programme schools. Rather, they are illustrative of the perceptions of headteachers, reading leaders and teachers in the selected programme schools. To mitigate this limitation, we have triangulated qualitative and quantitative findings whenever possible.

Lastly, caution should be exercised in generalising the results of this study. This trial was conducted with specific pupil age groups from non-representative schools throughout the UK while RWI and FS programming are offered for a wider age range of pupils and throughout the region.

Future research and publications

This evaluation suggests a few specific research questions that could be addressed by future research:

- What adaptations could be made to the programme or resource requirements (i.e., materials, budget, space) to support schools to implement the programme as intended?
- In what ways can support for the teachers be sustained over time?
- If the FS programme were implemented with higher fidelity, what would the effect be on reading and writing outcomes?
- What are effective strategies to support struggling readers in Years 5–8?

The results of this evaluation will also be published in comparative education journals as the implications are relevant for struggling English readers in different contexts worldwide, while underscoring the uniqueness of the UK context.

References

AGDEST. (2005). *Teaching reading: Report and recommendations.* Australian Government, Department of Education Science and Training, Commonwealth of Australia.

Chambers, B., Slavin, R. E., and Cheung, A. C. K. (2015). Literacy and language outcomes of balanced and developmental–constructivist approaches to early childhood education: A systematic review. *Center for Research and Reform in Education: Baltimore, MD.*.

Chenail, R. J. (2011). Interviewing the investigator: Strategies for addressing instrumentation and researcher bias concerns in qualitative research. *The Qualitative Report*, 16(1): 255.

Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory Into Practice*, 39(3): 124–130.

DFE. (2019a). National Curriculum assessments: Key Stage 2, 2019 (revised). Department for Education. Retrieved from https://www.gov.uk/government/statistics/national-curriculum-assessments-at-key-stage-2-2019-revised .

DFE. (2019b). Phonics screening check and Key Stage 1 assessments in England, 2019. Department for Education. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/851296/Phonics_screening_check_and_key_stage_1_assessments_in_England_2019.pdf

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

EEF. (2018). Statistical analysis guidance for EEF evaluations. Education Endowment Foundation.

Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. *Review of Educational Research*, 71(3): 393–447.

Gorard, S., Siddiqui, N., & See, B. H. (2015). *Fresh Start*. Education Endowment Foundation Efficacy Trial.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, 7(1): 6–10.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, 2(2): 127–160.

Hoover, W. A., & Tunmer, W. E. (2020). *The cognitive foundations of reading acquisition: A framework with applications connecting teaching and learning*. Switzerland: Springer, Cham.

Kim, Y-S. G. (2017). Why the simple view of reading is not simplistic: Unpacking component skills of reading using a direct and indirect effect model of reading (DIER). *Scientific Studies of Reading*, 4: 310–333.

Maxwell, J. A. (2005). *Qualitative research design: An interactive approach*. Thousand Oaks, CA: Sage Publishing.

NICHHD. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). National Institute of Child Health and Human Development, Washington, DC: U.S. Government Printing Office.

Rose, J. (2006).Independent review of the teaching of early reading: final report March 2006. Nottingham: DfES Publications.

Torgerson, C., Brooks, G., & Hall, J. (2006). *A systematic review of the research literature on the use of phonics in the teaching of reading and spelling*. Nottingham: DfES Publications.

White, H., & Phillips, D. (2012). Addressing attribution of cause and effect in small N impact evaluations: Towards an integrated framework. Working Paper 15. New Delhi, India: International Initiative for Impact Evaluation.

Appendix A: EEF cost rating

Figure A1: Cost rating

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

Appendix B: Security classification of trial findings

Read Write Inc

OUTCOME: NGRT

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

Threats to validity	Threat to internal validity?	Comments
Threat 1: Confounding	Low	Randomisation was conducted independently by the evaluation team. No pre-test was used, but groups were largely balanced at baseline across other key characteristics except for ethnicity and EAL status. Models cannot control for additional covariates as pupil characteristics are only available for pupils in Year 1 at randomisation.
Threat 2: Concurrent interventions	Low	Some evidence of the use of other interventions and strategies in the control group. There is little discussion as to whether there were additional programmes in use in the intervention group.
Threat 3: Experimental effects	Moderate	13 of 65 control schools received RWI training or materials during the trial, suggesting a fairly high degree of contamination.
Threat 4: Implementation fidelity	Moderate	Survey results suggest that most Reception pupils were taught RWI in the intended dosage, but only 64% of schools offered the recommended amount of RWI teaching to Y1 upwards across the two-year intervention period. IPE identifies several barriers to RWI delivery, including staffing and space needs and insufficient support for teaching assistants. In compliance analysis, 11 of 66 treatment schools were shown not to fully comply with RWI, but CACE estimates are similar to primary analysis results.
Threat 5: Missing data	Moderate	Very high level of missing data for RWI analysis, with some imbalance across groups (881 missing out of approx. 2534 pupils for treatment (35%) and 640 out of approx. 2380 (27%) for control). Results from multiple imputation models suggest the effect size may be slightly overestimated in the complete-case analysis (1 month vs. 0 months' progress), but confidence intervals span zero in both cases.
Threat 6: Measurement of outcomes	Moderate	Outcome assessments were administered by class teachers trained by external administrators. As such, they were not blind to treatment condition. Floor effects were also observed in NGRT scores, aligning with feedback from some teachers that the NGRT was too difficult for pupils.
Threat 7: Selective reporting	Low	Analysis follows pre-specified protocol and SAP.

- **Initial padlock score:** 2 padlocks – Randomised controlled trial with MDES at randomisation of 0.176 and attrition from primary analysis of 31%
- **Reason for adjustment for threats to validity**: 0 padlocks – Moderate threats to validity operate in opposite or unknown directions
- **Final padlock score:** Initial score adjusted for threats to validity = 2 padlocks

Fresh Start

OUTCOME: KS2 Reading

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

Threats to validity	Threat to internal validity?	Comments
Threat 1: Confounding	Low	Randomisation was conducted independently by the evaluation team. Minimal imbalance was observed at baseline across key characteristics, including pre-test KS1 score (effect size –0.007).
Threat 2: Concurrent interventions	Low	Some evidence of the use of other interventions and strategies in the control group. There is limited discussion of concurrent interventions in the treatment group in general and in schools that did not implement Fresh Start.
Threat 3: Experimental effects	Low	As usual practice data was collected at a single timepoint, conclusions cannot be drawn about changes in practice during the intervention period, although a proportion of control schools reported using Fresh Start resources at baseline.
Threat 4: Implementation fidelity	High	Very low levels of Fresh Start implementation fidelity: only 59% of schools reported implementing the intervention in the first year and 44% in the second. From compliance data, as many as 50 schools did not/ may not have fully delivered FS programming. Pupil-level data also suggests limited exposure to FS in treatment schools. Due to data limitations, compliance analysis could not be conducted.
Threat 5: Missing data	Moderate	The sample size at randomisation is not known with certainty, but overall missing data is estimated to be 16.7%, with some imbalance across groups (12.6% treatment, 20.1% control). Results from

		multiple imputation models are qualitatively similar to the primary analysis.
Threat 6: Measurement of outcomes	Low	Primary outcome measure uses KS2 national assessment of target year group. No floor or ceiling effects were observed.
Threat 7: Selective reporting	Low	Analysis follows pre-specified protocol and SAP.

- **Initial padlock score:** 4 padlocks – Randomised controlled trial with MDES at randomisation of 0.193 and overall missing data estimated at 16.7%
- **Reason for adjustment for threats to validity**: –1 padlocks – Very low implementation fidelity such that fewer than half of pupils were exposed to the intervention
- **Final padlock score:** initial score adjusted for threats to validity = 3 padlocks

Appendix C: Effect size estimation

Tables C1–C14 provide detailed information for the effect-size calculations for all analysis models of RWI and FS. The tables report the overall means, variance by intervention group and pooled variance for each outcome variable.

Primary analysis

Table C1: Effect-size estimation of RWI primary outcome

			Intervention group		Control group		
Outcome	**Unadjusted differences in means**	**Adjusted differences in means**	**n (missing)**	**Variance of outcome**	**n (missing)**	**Variance of outcome**	**Pooled variance**
NGRT – SAS	0.20	0.67	1,653 (881)	214.76	1,740 (640)	205.61	210.02

Table C2: Effect-size estimation of FS primary outcome

			Intervention group		Control group		
Outcome	**Unadjusted differences in means**	**Adjusted differences in means**	**n (missing)**	**Variance of outcome**	**n (missing)**	**Variance of outcome**	**Pooled variance**
KS2 reading marks – Year 6	–2.57	–2.04	637 (47)	86.20	561 (54)	83.66	86.73
KS2 reading marks – Year 5	–0.53	–0.33	579 (91)	105.71	541 (63)	95.63	100.92
KS2 reading marks – combined	–1.64	–1.24	1,216 (138)	100.26	1,102 (117)	91.06	96.68

Secondary analysis

Table C3: Effect-size estimation of RWI secondary outcomes

			Intervention group		Control group		
Outcome	**Unadjusted differences in means**	**Adjusted differences in means**	**n (missing)**	**Variance of outcome**	**n (missing)**	**Variance of outcome**	**Pooled variance**
KS1 writing passed – Year 3	–0.03	–0.03	2,403 (34)	0.23	2,320 (31)	0.22	0.23
KS1 writing passed – Year 2	–0.03	–0.02	2,496 (38)	0.22	2,343 (37)	0.21	0.22
KS1 writing passed – combined	–0.03	–0.02	4,899 (72)	0.23	4,663 (68)	0.22	0.22
Phonics screening score – Year 2	–0.01	0.13	2,492 (42)	78.25	2,355 (25)	75.33	76.81
Phonics screening score – Year 1	1.28	1.38	1,710 (22)	47.86	1,989 (0)	70.67	60.51
Phonics screening score – combined	0.49	0.65	4,202 (64)	66.63	4,344 (25)	73.24	70.04
Passed phonics – Year 2	–0.00	0.00	2,492 (42)	0.15	2,355 (25)	0.15	0.15
Passed phonics – Year 1	0.04	0.05	1,710 (22)	0.10	1,989 (0)	0.13	0.11
Passed phonics – combined	0.02	0.02	4,202 (64)	0.13	4,344 (25)	0.14	0.13

Table C4: Effect-size estimation of FS secondary outcome

			Intervention group		Control group		
Outcome	**Unadjusted differences in means**	**Adjusted differences in means**	**n (missing)**	**Variance of outcome**	**n (missing)**	**Variance of outcome**	**Pooled variance**
KS2 writing score — combined	–0.03	0.17	1,324 (30)	88.76	1,188 (31)	85.14	87.01

Analysis in the presence of non-compliance

Table C5: Effect-size estimation of RWI CACE analysis

			Intervention group		Control group		
Outcome	**Unadjusted differences in means**	**Adjusted differences in means**	**n (missing)**	**Variance of outcome**	**n (missing)**	**Variance of outcome**	**Pooled variance**
NGRT – SAS	0.27	0.89	1,846 (705)	212.81	1,547 (816)	206.46	210.02
KS1 writing passed – Year 3	–0.06	–0.05	2,482 (39)	0.23	2,241 (26)	0.23	0.23
KS1 writing passed – Year 2	–0.05	–0.04	2,490 (61)	0.22	2,349 (14)	0.21	0.22
KS1 writing passed – combined	–0.05	–0.04	4,972 (100)	0.22	4,590 (40)	0.22	0.22
Phonics screening score – Year 2	–0.01	0.23	2,484 (67)	73.78	2,363 (0)	80.00	76.81
Phonics screening score – Year 1	1.62	1.71	2,014 (20)	48.15	1,685 (2)	74.16	60.51
Phonics screening score – combined	0.74	0.97	4,498 (87)	62.96	4,048 (2)	77.59	70.04
Passed phonics – Year 2	–0.00	0.01	2,484 (67)	0.15	2,363 (0)	0.15	0.15
Passed phonics – Year 1	0.05	0.06	2,014 (20)	0.10	1,685 (2)	0.13	0.11
Passed phonics – combined	0.02	0.03	4,498 (87)	0.13	4,048 (2)	0.14	0.13

Missing data analysis

Table C7: Effect-size estimation of RWI missing data analysis

			Intervention group		Control group		
Outcome	**Unadjusted differences in means**	**Adjusted differences in means**	**n (missing)**	**Variance of outcome**	**n (missing)**	**Variance of outcome**	**Pooled variance**
NGRT – SAS	–0.09	0.55	2,534 (0)	231.90	2,380 (0)	298.69	277.98
KS1 writing passed – Year 3	–0.03	–0.03	2,437 (0)	0.24	2,351 (0)	0.23	0.23
KS1 writing passed – Year 2	–0.03	–0.02	2,534 (0)	0.23	2,380 (0)	0.21	0.22
KS1 writing passed – combined	–0.03	–0.02	4,971 (0)	0.23	4,731 (0)	0.23	0.23
Phonics screening score – Year 2	–0.04	0.11	2,534 (0)	80.73	2,380 (0)	76.66	79.66
Phonics screening score – Year 1	1.28	1.38	1,732 (0)	47.86	1,989 (0)	70.67	60.51
Phonics screening score – combined	0.46	0.62	4,266 (0)	68.00	4,369 (0)	73.99	71.77

Table C8: Effect-size estimation of FS missing data analysis

			Intervention group		Control group		
Outcome	**Unadjusted differences in means**	**Adjusted differences in means**	**n (missing)**	**Variance of outcome**	**n (missing)**	**Variance of outcome**	**Pooled variance**
KS2 reading marks – Year 6	–2.42	–1.85	684 (0)	96.90	615 (0)	95.30	101.49
KS2 reading marks – Year 5	–0.37	–0.08	670 (0)	122.26	604 (0)	108.74	115.97
KS2 reading marks – combined	–1.40	–0.99	1,354 (0)	120.94	1,219 (0)	102.41	115.70
KS2 writing score — combined	–0.05	0.13	1,354 (0)	88.39	1,219 (0)	87.94	87.91

Subgroup analyses

Table C9: Effect-size estimation of RWI outcomes for FSM pupils

			Intervention group		Control group		
Outcome	**Unadjusted differences in means**	**Adjusted differences in means**	**n (missing)**	**Variance of outcome**	**n (missing)**	**Variance of outcome**	**Pooled variance**
NGRT – SAS	1.50	3.30	456 (279)	211.55	341 (159)	221.14	215.93
KS1 writing passed – Year 3	0.02	0.04	810 (18)	0.25	606 (5)	0.25	0.25
KS1 writing passed – Year 2	–0.03	–0.00	733 (2)	0.25	485 (15)	0.25	0.25
KS1 writing passed – combined	0.00	0.02	1,543 (20)	0.25	1,091 (20)	0.25	0.25
Phonics screening score – Year 2	–0.02	0.60	724 (11)	110.76	482 (18)	108.02	109.58
Passed phonics – Year 2	–0.01	0.01	724 (11)	0.20	482 (18)	0.19	01.9

Table C10: Effect-size estimation of FS outcomes for FSM pupils

			Intervention group		Control group		
Outcome	**Unadjusted differences in means**	**Adjusted differences in means**	**n (missing)**	**Variance of outcome**	**n (missing)**	**Variance of outcome**	**Pooled variance**
KS2 reading marks – Year 6	–2.44	–2.45	344 (32)	85.71	255 (34)	85.31	86.86
KS2 reading marks – Year 5	–1.91	–1.40	281 (54)	114.79	263 (39)	103.40	110.00
KS2 reading marks – combined	–2.38	–2.05	625 (86)	101.77	518 (73)	96.62	100.75
KS2 writing score – combined	–0.00	0.33	696 (15)	96.17	578 (13)	89.14	92.91

Table C11: Effect-size estimation of RWI outcomes for female pupils

			Intervention group		Control group		
Outcome	**Unadjusted differences in means**	**Adjusted differences in means**	**n (missing)**	**Variance of outcome**	**n (missing)**	**Variance of outcome**	**Pooled variance**
NGRT – SAS	1.24	1.50	825 (442)	194.21	871 (295)	197.71	194.74
KS1 writing passed – Year 3	–0.03	–0.02	1,202 (16)	0.21	1,147 (5)	0.20	0.21
KS1 writing passed – Year 2	–0.01	–0.00	1,241 (26)	0.19	1,149 (17)	0.18	0.19
KS1 writing passed – combined	–0.02	–0.01	2,443 (42)	0.20	2,296 (22)	0.19	0.20
Phonics screening score – Year 2	0.44	0.48	1,239 (28)	48.86	1,161 (5)	54.72	51.72
Passed phonics – Year 2	0.02	0.02	1,239 (28)	0.11	1,161 (5)	0.12	0.12

Table C12: Effect-size estimation of FS outcomes for female pupils

			Intervention group		Control group		
Outcome	**Unadjusted differences in means**	**Adjusted differences in means**	**n (missing)**	**Variance of outcome**	**n (missing)**	**Variance of outcome**	**Pooled variance**
KS2 reading marks – Year 6	–2.64	–2.03	254 (20)	77.81	242 (16)	84.02	82.42
KS2 reading marks – Year 5	–1.17	–0.62	242 (33)	106.60	221 (27)	92.58	100.04
KS2 reading marks – combined	–1.89	–1.15	496 (53)	96.65	463 (43)	90.12	94.31
KS2 writing score – combined	0.46	0.81	536 (13)	74.13	488 (18)	80.11	76.96

Additional analyses and robustness checks

Table C13: Effect-size estimation of RWI additional analysis

			Intervention group		Control group		
Outcome	**Unadjusted differences in means**	**Adjusted differences in means**	**n (missing)**	**Variance of outcome**	**N (missing)**	**Variance of outcome**	**Pooled variance**
NGRT – SAS	0.25	0.73	1,653 (881)	214.76	1,759 (742)	205.70	210.04
KS1 writing passed – Year 3	–0.03	–0.02	2,403 (34)	0.23	2,443 (33)	0.23	0.23
KS1 writing passed – Year 2	–0.03	–0.02	2,492 (42)	0.22	2,460 (41)	0.21	0.22
KS1 writing passed – combined	–0.03	–0.02	4,899 (72)	0.23	4,903 (74)	0.22	0.22
Phonics screening score – Year 2	–0.01	0.17	2,492 (42)	78.25	2,475 (26)	74.39	76.31
Phonics screening score – Year 1	1.25	1.36	1,710 (22)	47.86	2,044 (15)	70.61	60.62
Phonics screening score – combined	0.48	0.66	4,202 (64)	66.63	4,519 (41)	72.73	69.84
Passed phonics – Year 2	–0.00	0.00	2,492 (42)	0.15	2,475 (26)	0.15	0.15
Passed phonics – Year 1	0.04	0.05	1,710 (22)	0.10	2,044 (15)	0.13	0.11
Passed phonics – Combined	0.01	0.02	4,202 (64)	0.13	4,519 (41)	0.14	0.13

Table C14: Effect-size estimation of FS additional analysis

			Intervention group		Control group		
Outcome	**Unadjusted differences in means**	**Adjusted differences in means**	**N (missing)**	**Variance of outcome**	**n (missing)**	**Variance of outcome**	**Pooled variance**
KS2 reading marks – Year 6	–2.66	–2.28	637 (47)	86.20	585 (54)	85.93	87.87
KS2 reading marks – Year 5	–0.63	–0.45	579 (91)	105.71	573 (65)	100.99	103.49
KS2 reading marks – combined	–1.75	–1.44	1,216 (138)	100.26	1,158 (119)	94.99	98.58
KS2 writing score — combined	–0.00	0.15	1,324 (30)	88.76	1,245 (32)	86.59	87.68

Appendix D: Analysis of missing data

To analyse the patterns of missing data, we focused on our primary outcome measures for both RWI and FS, and investigated whether a missing outcome was related to pupil ethnicity, gender, FSM status or EAL status. In addition, we investigated whether the relationships of these measures with a missing outcome differ by intervention and control schools. To conduct this investigation, we estimated a series of logit regression models and used the likelihood ration test (LRT) chi-squared to examine our hypotheses.

Table D1 reports the p-values of the LRTs for pupils' gender, FSM status and EAL status. Tests including covariates are also reported. The covariates include pupils' gender, FSM status, EAL status and the stratification variables used for randomisation. The characteristics tested are excluded from the covariate vector for that particular analysis. The sign on the coefficients and LRT results indicates significant relationships: missingness in the outcome variable is positively associated with pupil FSM status for both RWI and FS and is positively associated with EAL status for FS pupils. In other words, FSM pupils are more likely to be missing data for both programmes and EAL pupils are more likely to be missing it for FS.

Table D1: LRT of models predicting missing pupil primary outcomes

Test conducted	Likelihood ratio test p-value
RWI	
Pupil FSM vs. Null	0.0234
Pupil FSM with covariates vs. only covariates	0.0342
Pupil EAL vs. Null	0.4109
Pupil EAL with covariates vs. only covariates	0.1028
Female vs. Null	0.1880
Female with covariates vs. only covariates	0.2790
FS	
Pupil FSM vs. Null	0.0014
Pupil FSM with covariates vs. only covariates	0.0025
Pupil EAL vs. Null	0.8102
Pupil EAL with covariates vs. only covariates	0.0279
Female vs. Null	0.2718
Female with covariates vs. only covariates	0.4692

Table D2 reports p-values of the LRTs examining whether the relationship of missing primary outcomes with pupil characteristics differ by intervention and control schools. In this analysis, all models additionally include the intervention indicator as a covariate. The contrast is between a model including an interaction between the intervention indicator and the variable of interest compared to the model without the intervention interaction. If the interaction is statistically significant, that indicates a different relationship in intervention and control schools. The LRT results for this study find evidence of significant differences when the interaction term is included for RWI but not for FS. This result indicates that missingness in the outcome variable is differentially related to pupil characteristics across intervention and control schools for RWI but not for FS.

Table D2: LRT of models predicting missing pupil primary outcomes with interactions

Test conducted	Likelihood ratio test p-value
RWI	
Pupil FSM × Intervention vs. Pupil FSM	0.0000
Pupil FSM × Intervention with covariates vs. Pupil FSM with covariates	0.0000
Pupil EAL × Intervention vs. Pupil EAL	0.0000
Pupil EAL × Intervention with covariates vs. Pupil EAL with covariates	0.0000
Female × Intervention vs. Female	0.0000
Female × Intervention with covariates vs. Female with covariates	0.0000
FS	
Pupil FSM × Intervention vs. Pupil FSM	0.7181
Pupil FSM × Intervention with covariates vs. Pupil FSM with covariates	0.6889
Pupil EAL × Intervention vs. Pupil EAL	0.5502
Pupil EAL × Intervention with covariates vs. Pupil EAL with covariates	0.7162
Female × Intervention vs. Female	0.6030
Female × Intervention with covariates vs. Female with covariates	0.7236

In summary, we found that missing primary outcomes was related to pupil FSM status for both RWI and FS pupils. Pupil FSM status was differentially related in intervention and control schools for RWI only. Given the level of missing data in the outcome variables and the relationship with key covariates, we employed multiple imputation to fill in missing values and estimated our primary analysis models on these data to check the robustness of results.

Appendix E: Example analysis code

All analysis was conducted using Stata 14 (StataCorp., 2017). The primary models for RWI NGRT scores (sas) were estimated using the following Stata command:

```
reg sas intervention i.strata if itt==1 & readwriteinc==1, cluster(schoolid)
```

The RWI model includes the categorical randomisation strata as covariates. The main models for FS KS2 reading scores (ks_readscore) were estimated using the following Stata commend:

```
reg ks2_readscore intervention i.strata if itt==1 & freshstart==1, cluster(schoolid)
```

The FS models included pupils' standardised KS1 reading scores as a covariate along with the strata. All models clustered standard errors at the school level (schoolid).

Appendix F: Memorandums of Understanding and Consent Forms

F. 1 EEF funds large trial of Read, Write, Inc. and Fresh Start

Information sheet for Headteachers

Read, Write, Inc and Fresh Start

RWI teaches young children to read and write, through a structured and systematic approach to teaching literacy. It is used by more than a quarter of the UK's primary schools. Fresh start is a similar programme, but is a catch up programme for those children that are struggling to read at the end of primary and the start of secondary. The reading materials used in Fresh Start are age-appropriate and older children do not feel they are reading books designed for much younger children.

Both programmes start with the systematic teaching of phonemes and graphemes, carefully matching the sounds that children have been taught with books that contain only those sounds. The programmes then move on to improving reading fluency and comprehension.

Schools receive two whole-school training days, with the training covering the principles and teaching techniques required to teach both RWI and FS. In addition, schools get between three to six developmental days per year, depending on their starting point for teaching reading within the school. Schools are encouraged to run sessions for parents, these teach parents the sounds that children will be learning and shows them how to use the books that children take home.

Why is the EEF funding a trial of these programmes?

Fresh Start has been tested through a previous EEF efficacy trial involving ten secondary schools and 433 Year 7 pupils. The study randomised pupils within each of the schools and focused particularly on pupils who did not achieve a level 4 on their KS2 SATs. The project found a positive impact on all pupils of an additional three months' progress over the course of an academic year. RWI uses identical techniques with younger children with the aim of preventing reading difficulties so that later interventions aren't required or so that the number of children who need them is minimised.

Who can participate in this trial?

All Primary schools in England can participate in this trial, if they have not yet received full training from Ruth Miskin Training to deliver Read Write Inc. Phonics and Fresh start.

Primary schools in the North East are particularly being encouraged to take part in this trial as part of the EEF's North East Primary Literacy Campaign.

Recruitment is now open for 120 Primary schools to participate in this trial. Of the schools who agree to participate 60 will be randomly allocated to receive Read Write Inc. Phonics and Fresh Start. The other 60 schools will act as a business as usual control group, continuing with their normal teaching of reading and support for struggling readers and receiving a financial incentive for taking part in the project.

When and how can schools sign up to participate in the trial?

When? The recruitment process has started and schools will be recruited on a first come first served basis, so speed is recommended if a school wishes to participate in this trial.

How? Schools who wish to participate should contact Dr Maria Cockerill at Queen's University, Belfast: **cockerill_maria@yahoo.co.uk** . Schools who sign up to participate in the trial will be sent a Memorandum of Agreement to complete. This agreement clarifies what schools will receive and what they will need to do during the trial.

What will schools who participate in the trial do?

Schools who sign up to participate in the trial will either be selected to deliver the programme Read Write Inc. Phonics and Fresh Start between September 2016 and July 20l18, or to act as a control school until August 2018.

How will schools be selected to be a programme or control school?

Schools will be randomly selected to participate as either a Programme School or a Control School. As selection is randomly undertaken schools have an equal chance of being selected to either group. **Programme Schools**, who will deliver Read Write Inc. Phonics and Fresh Start between September 2016 and July 2018. **Control Schools**, who will continue until Aug 18 with their normal teaching of reading and support for struggling readers.

When will schools know if they have been selected to be a Programme School or Control School?

Randomisation of schools into programme or control will occur in phases, monthly, beginning on April 19th, May 19th, June 19th and so on. Schools who sign up to participate in the trial during this period will be informed if they are programme or control school following the randomisation dates outlined here.

What will Programme Schools or Control Schools receive and commit to do?

Programme Schools will:

- Agree to deliver the programme in KS1 and KS2 between September 2016 and July 2018.
- Receive free staff training and ongoing support and development days from Ruth Miskin Training during the trial.
- Be eligible for a 20% discount from Oxford University Press to purchase all Read Write Inc. Phonics and Fresh Start resources during the trial. These can be ordered from April 25th 2016.

Control Schools will:

  - Agree to wait to deliver Read Write Inc. Phonics and Fresh Start in their school until August 2018.
  - Receive £3000 for participating as control schools between September 2016 and July 2018.
- Be eligible for a 20% discount from Oxford University Press to purchase all Read Write Inc. Phonics and Fresh Start resources from July 2018.

All Schools will:

  - Distribute opt-out consent letters to parents of pupils who will be in Year 2 and Year 6 in January 2018.
  - Be expected to provide access to reading test results from the Granada Learning NGRT from the School Management System.
- Provide basic school level information that will be used to help ensure balance between control and intervention school sets.

If you are interested please complete the information sheet overleaf and also sign the consent to participate form.

Your decision to participate/not participate in this trial will not impact upon your (or indeed those of anyone in your school) existing or future relationships with Queen's University Belfast.

If you have any questions please don't hesitate to contact Dr Nicole Craig, 6 College Green, School of Education, Queen's University Belfast, Belfast BT71NN, or phone on 028 9097 5017 or via email at **n.craig@qub.ac.uk**. Or Dr Maria Cockerill at Queen's University, Belfast: **cockerill_maria@yahoo.co.uk.**

If you have a complaint regarding the work then contact Allen Thurston at Queen's University Belfast on 07964916186 or via **a.thurston@qub.ac.uk**

EEF trial of *Read Write Inc.* Phonics and Fresh Start

September 2016 – July 2018

SCHOOL REGISTRATION AND CONSENT TO TRIAL	
School name	
School address	
School Telephone	
LA and County	
Headteacher name	
Headteacher contact email	
School admin email	
School Ofsted rating	
FSM %	
Ever FSM %	
Number of pupils on roll	
% No. of Boys % No. of Girls	

1, 2 or 3 class entry school	
2015 KS1 Reading results	
2015 KS2 Reading results	
% achieving level 4 or above in reading, writing and maths	
% achieving level 4b or above in Reading and maths tests and level 4 or above in Writing Teacher Assessed	
Gap between disadvantaged pupils and others 2015	
% EAL pupils who achieve L4 or above in reading, writing and maths	
Any other information you consider relevant	

I consent to my school being involved with the Read, Write Inc/Fresh Start project. I have read and understand the information sheets and what will be required of schools. I understand that the school can withdraw participation at any time.

Your Name ……..

Your Signature ……..……

Headteacher/Other (please specify)……

Date ……..……

F. 2 Parental Information Form for Read, Write Inc

Dear Parent/Guardian,

Your child has been invited to continue to take part in a research study. Before you decide whether you would like them to continue to take part, we would like to inform you about why the next stage of research is being conducted and what it will involve. We would appreciate it if you could please take a few moments to read the following information carefully.

Your child's school is involved in a two-year programme called *Read, Write Inc/Fresh Start*. It is a partnership between Ruth Miskin, a team from the Centre for Evidence and Social Innovation at Queen's University Belfast, Northern Ireland and your child's school. The programme is being evaluated by a team from the American Institute for Research. The programme aims to improve reading.

As part of this programme we will require access data from the National Pupil Database for the children at the end of 2018. Therefore, this letter and consent form is asking you to tell us if you do NOT allow us to access your child's data from the National Pupil Database.

We require these scores to establish how successful the Read, Write Inc/Fresh Start programme has been in improving children's educational attainment. We are evaluating whether the programme has benefited children (compared to children who have not had the programme), and not how well your child is doing in school. Data will be held securely for a minimum of five years by Queen's University Belfast and the funder of the research, The Education Endowment Foundation and then securely destroyed. Electronic data will be held at a secure site on an encrypted computer. Your child's name will not be collected from the school.

As participation is voluntary, your child is free to withdraw from the study at any time up until the point that the data set is made anonymous, after that we will be unable to withdraw them as we will not be able to identify them in the data set. Research reports will be the main products of the research and these will specifically attempt to determine the effects Read, Write Inc/Fresh Start has on literacy outcomes. Any reports will not mention individual names or even the name of schools that participated in the project.

A decision to participate (or not) will not affect your relationship with your child's school or with Queen's University Belfast. The project has ethical approval from Queen's University Belfast School of Education ethics Committee.

If you **do NOT wish** us to access your child's data from the National Pupil Database, please complete the details on the consent form below, sign it, and return to your child's class teacher. If you are happy for us to access this data no further action from you is required.

Since we started this project new rules have come into effect regarding use and processing of data. These rules are called GDPR. You/your child have rights to privacy under these rules that include:

The right to be informed of who is accessing your/your child's data and why (this letter outlines what data we are accessing and why)

You have a right **of** access to your/your child's data (you should contact your school should you wish to access your child's national pupil database extract)

You have right to rectification if you feel any data is incorrect (contact us or the school to take this forward)

You have a right to erasure/to be forgotten (if you contact us saying you wish to withdraw from the project we will securely erase your child's record from our data permanently)

You also have a right to restrict processing (again by requesting consent is withdrawn we would not process any data about your child)

Other rights include the right to data portability, right to object **and rights in** relation to automated decision making **and** profiling. However, we do not consider that these aspects of data processing are relevant to this project.

If you have any questions please don't hesitate to contact Dr Nina O'Neill, 6 College Green, School of Education, Queen's University Belfast, Belfast BT71NN, or phone on 028 9097 5017 or via email at **nina.oneill@qub.ac.uk**

If you have a complaint regarding the work then contact Allen Thurston at Queen's University Belfast on 07964916186 or via **a.thurston@qub.ac.uk**

F.3 Parental Opt-out Consent Form for Read, Write Inc

Please return this form to your child's class teacher if you **are NOT** willing for your child's Phonics Check scores to be accessed from the National Pupil Database.

I **do NOT** give permission for my child's Phonics Check to be accessed in the National Pupil Database as part of the Read, Write Inc/Fresh Start research project.

Your Child's Name ……..……

Your Name ……..

Your Signature ……..……

Parent / Guardian (delete as appropriate)

Date ……..……

Read, Write Inc/Fresh Start is provided by Ruth Miskin and funded by The Education Endowment Foundation

F.4 Parental Information Form for Read, Write Inc – Updated

Dear Parent/Guardian,

Your child has been invited to continue to take part in a research study. Before you decide whether you would like them to continue to take part, we would like to inform you about why the next stage of research is being conducted and what it will involve. We would appreciate it if you could please take a few moments to read the following information carefully.

Your child's school is involved in a two-year programme called *Read, Write Inc/Fresh Start*. It is a partnership between Ruth Miskin, a team from the Centre for Evidence and Social Innovation at Queen's University Belfast, Northern Ireland and your child's school. The programme is being evaluated by a team from the American Institute for Research. The programme aims to improve reading.

As part of this programme we will require access data from the National Pupil Database for the children at the end of 2018. Therefore, this letter and consent form is asking you to tell us if you do NOT allow us to access your child's data from the National Pupil Database.

We require these scores to establish how successful the Read, Write Inc/Fresh Start programme has been in improving children's educational attainment. We are evaluating whether the programme has benefited children (compared to children who have not had the programme), and not how well your child is doing in school. Data will be held securely for a minimum of five years by Queen's University Belfast and the funder of the research, The Education Endowment Foundation and then securely destroyed. Electronic data will be held at a secure site on an encrypted computer. Your child's name will not be collected from the school.

As participation is voluntary, your child is free to withdraw from the study at any time up until the point that the data set is made anonymous, after that we will be unable to withdraw them as we will not be able to identify them in the data set. Research reports will be the main products of the research and these will specifically attempt to determine the effects Read, Write Inc/Fresh Start has on literacy outcomes. Any reports will not mention individual names or even the name of schools that participated in the project.

A decision to participate (or not) will not affect your relationship with your child's school or with Queen's University Belfast. The project has ethical approval from Queen's University Belfast School of Education ethics Committee.

If you **do NOT wish** us to access your child's data from the National Pupil Database, please complete the details on the consent form below, sign it, and return to your child's class teacher. If you are happy for us to access this data no further action from you is required.

Since we started this project new rules have come into effect regarding use and processing of data. These rules are called GDPR. You/your child have rights to privacy under these rules that include:

The right to be informed of who is accessing your/your child's data and why (this letter outlines what data we are accessing and why)

You have a right **of** access to your/your child's data (you should contact your school should you wish to access your child's national pupil database extract)

You have right to rectification if you feel any data is incorrect (contact us or the school to take this forward)

You have a right to erasure/to be forgotten (if you contact us saying you wish to withdraw from the project we will securely erase your child's record from our data permanently)

You also have a right to restrict processing (again by requesting consent is withdrawn we would not process any data about your child)

Other rights include the right to data portability, right to object **and rights in** relation to automated decision making **and** profiling. However, we do not consider that these aspects of data processing are relevant to this project.

If you have any questions please don't hesitate to contact Dr Nina O'Neill, 6 College Green, School of Education, Queen's University Belfast, Belfast BT71NN, or phone on 028 9097 5017 or via email at **nina.oneill@qub.ac.uk**

If you have a complaint regarding the work then contact Allen Thurston at Queen's University Belfast on 07964916186 or via **a.thurston@qub.ac.uk**

F.5 Parental Opt-out Consent Form for Read, Write Inc – Updated

Please return this form to your child's class teacher if you **are NOT** willing for your child's Phonics Check scores to be accessed from the National Pupil Database.

I **do NOT** give permission for my child's Phonics Check to be accessed in the National Pupil Database as part of the Read, Write Inc/Fresh Start research project.

Your Child's Name ……..……

Your Name ……..

Your Signature ……..……

Parent / Guardian (delete as appropriate)

Date ……..……

Read, Write Inc/Fresh Start is provided by Ruth Miskin and funded by The Education Endowment Foundation

F.6 Parental Information Form for Fresh Start

Dear Parent/Guardian,

Your child has been invited to continue to take part in a research study. Before you decide whether you would like them to continue to take part, we would like to inform you about why the next stage of research is being conducted and what it will involve. We would appreciate it if you could please take a few moments to read the following information carefully.

Your child's school is involved in a two-year programme called *Read, Write Inc/Fresh Start*. It is a partnership between Ruth Miskin, a team from the Centre for Effective Education at Queen's University Belfast, Northern Ireland and your child's school. The programme is being evaluated by a team from the American Institute for Research. The programme aims to improve reading.

As part of this programme we will require access to the Key Stage 1, Key Stage 2 and standardised reading test scores from your child's school and the National Pupil Database for the children at the end of Year 2 and Year 6 respectively. **Therefore, this letter and consent form is asking you to tell us if you <u>do NOT allow</u> us to access Key Stage scores from the National Pupil Database and Reading Test scores from the school.**

We require these scores to establish how successful the Read, Write Inc/Fresh Start programme has been in improving children's educational attainment. We are evaluating whether the programme has benefited children (compared to children who have not had the programme), and not how well your child is doing in school. Data will be held securely for a minimum of five years by Queen's University Belfast and the funder of the research, The Educational Endowment Foundation and then securely destroyed. Electronic data will be held at a secure site on an encrypted computer. Your child's name will not be stored.

As participation is voluntary, your child is free to withdraw from the study at any time up until the point that the data set is made anonymous, after that we will be unable to withdraw them as we will not be able to identify them in the data set. Research reports will be the main products of the research and these will specifically attempt to determine the effects Read, Write Inc/Fresh Start has on literacy outcomes. Any reports will not mention individual names or even the name of schools that participated in the project.

A decision to participate (or not) will not affect your relationship with your child's school or with Queen's University Belfast. The project has ethical approval from Queen's University Belfast School of Education ethics Committee.

If you **do NOT wish** us to access your child's Key Stage data from the National Pupil Database, please complete the details on the consent form below, sign it, and return to your child's class teacher. If you are happy for us to access this data no further action from you is required.

If you have any questions please don't hesitate to contact Dr Nicole Craig, 6 College Green, School of Education, Queen's University Belfast, Belfast BT71NN, or phone on 028 9097 5017 or via email at **n.craig@qub.ac.uk**.

If you have a complaint regarding the work then contact Allen Thurston at Queen's University Belfast on 07964916186 or via **a.thurston@qub.ac.uk**

F.7 Parental Opt-out Consent Form for Fresh Start

Please return this form to your child's class teacher if you **are NOT** willing for your child's Key Stage 1/Key Stage 2 scores to be accessed from the National Pupil Database/ Reading Test scores to be accessed via the school.

I **do NOT** give permission for my child's Key Stage scores to be accessed in the National Pupil Database as part of the Read, Write Inc/Fresh Start research project.

Your Child's Name ……..……

Your Name ……..……..

Your Signature ……..……..……

Parent / Guardian (delete as appropriate)

Date ……..……

Read, Write Inc/Fresh Start is provided by Ruth Miskin and funded by The Education Endowment Foundation

F.8 Parental Information Form for Fresh Start – Updated

Dear Parent/Guardian,

Your child has been invited to continue to take part in a research study. Before you decide whether you would like them to continue to take part, we would like to inform you about why the next stage of research is being conducted and what it will involve. We would appreciate it if you could please take a few moments to read the following information carefully.

Your child's school is involved in a two-year programme called *Fresh Start*. It is a partnership between Ruth Miskin, a team from the Centre for Evidence and Social Innovation at Queen's University Belfast, Northern Ireland and your child's school. The programme is being evaluated by a team from the American Institute for Research. The programme aims to improve reading.

As part of this programme we will require access KS2 assessment scores from the National Pupil Database for the children at the end of Year 6 in 2017/18. **Therefore, this letter and consent form is asking you to tell us if you DO NOT allow us to access the KS2 data from the National Pupil Database for your child.**

We require these scores to establish how successful the Fresh Start programme has been in improving children's educational attainment. We are evaluating whether the programme has benefited children (compared to children who have not had the programme), and not how well your child is doing in school. Any data Queen's University Belfast have will be held securely for a minimum of five years by Queen's University Belfast and the funder of the research, The Education Endowment Foundation and securely destroyed. Electronic data will be held at a secure site on an encrypted computer. Your child's name will not be collected from the school.

As participation is voluntary, your child is free to withdraw from the study at any time up until the point that the data set is made anonymous, after that we will be unable to withdraw them as we will not be able to identify them in the data set. Research reports will be the main products of the research and these will specifically attempt to determine the effects Fresh Start has on literacy outcomes. Any reports will not mention individual names or even the name of schools that participated in the project.

A decision to participate (or not) will not affect your relationship with your child's school or with Queen's University Belfast. The project has ethical approval from Queen's University Belfast School of Education Ethics Committee. This consent is ethical consent for us to proceed. We will use Article 6(1)e of the GDPR as the lawful basis for processing personal data as part of this project. This is generally known as the "public task" basis. We have reviewed current ICO guidance available here: **https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/public-task/**, and have determined that this research forms part of Queen's University's performance of a task in the public interest, as one of its core purposes provided for in its Charter and Statutes. The public interest will be in determining whether Fresh Start improves learning in schools.

If you **do NOT wish** us to access your child's KS2 data from the National Pupil Database, please complete the details on the consent form below, sign it, and return to your child's class teacher. If you are happy for us to access this data no further action from you is required.

Since we started this project new rules have come into effect regarding privacy, and the use and processing of data. These rules are called GDPR. You/your child have rights to privacy under these rules that include:

- The right to be informed of who is accessing your/your child's data and why (this letter outlines what data we are accessing and why)
- You have a right **of** access to your/your child's data (you should contact your school should you wish to access your child's national pupil database extract)
- You have right to rectification if you feel any data is incorrect (contact us or the school to take this forward)
- You have a right to erasure/to be forgotten (if you contact us saying you wish to withdraw from the project we will securely erase your child's record from our data permanently)
- You also have a right to restrict processing (again by requesting consent is withdrawn we would not process any data about your child)

- Other rights include the right to data portability, right to object **and rights in** relation to automated decision making **and** profiling. However, we do not consider that these aspects of data processing are relevant to this project.

If you have any questions please don't hesitate to contact Dr Joanne O'Keeffe, 6 College Green, School of Social Sciences, Education & Social Work, Queen's University Belfast, Belfast BT71NN, or phone on 028 9097 5017 or via email at **j.okeeffe@qub.ac.uk**

If you have a complaint regarding the work then contact Allen Thurston at Queen's University Belfast on 07964916186 or via **a.thurston@qub.ac.uk**

F.9 Parental Opt-out Consent Form for Fresh Start – Updated

Please return this form to your child's class teacher if you **are NOT** willing for your child's KS2 scores to be accessed from the National Pupil Database.

I **DO NOT** give permission for my child's KS2 to be accessed in the National Pupil Database as part of the Fresh Start research project.

Your Child's Name ……..……

Your Name ……

Your Signature ……..……

Parent / Guardian (delete as appropriate)

Date ……..……

Fresh Start is provided by Ruth Miskin and funded by The Education Endowment Foundation

Appendix G: Explanation of Fresh Start Fidelity

For FS programming, RMT collected data from programme schools for pupils who were originally identified as eligible for FS during study recruitment and randomisation phases. The UPNs of these eligible studies were obtained and shared with QUB and AIR researchers to track the pupils over the course of the trial and to match pupil outcomes with NPD data for the analyses. RMT intended to track these pupils to see whether or not they received the FS intervention as intended in schools assigned to treatment. Of the original 1,212 pupils in treatment schools identified as eligible for the intervention, RMT was only able to track the status of FS receipt for 932 pupils. Of these 932 pupils, they discovered 719 never received FS programming, while 352 did receive FS programming. They were unable to collect data for 205 pupils originally eligible for FS. Table G1, below, presents the breakdown of the information RMT received from schools about the status of children receiving FS programming or not, and how they were ultimately categorised as having received or not receiving the intervention. As shown, FS programming was not implemented with high fidelity throughout schools assigned to receive the intervention. Further, there is wide variation in the amount of FS programming received – with receipt of any of the programming being categorised as receiving the intervention – within the group of pupils who were noted as having received FS programming. Due to inability to match pupil records with NPD outcomes, we were unable to conduct a CACE analysis for FS.

Table G1: Categorisation of receipt of FS for FS-eligible pupils

FS status	Definition	Pupils	Category
CME – never received Fresh Start	Children Missing in Education – never received intervention	12	Never received
CME – some Fresh Start	Children Missing in Education – sporadic intervention	1	Received
Left never received	Left before intervention started	46	Never received
Left some received	Left while receiving Fresh Start	6	Received
No Fresh Start received	Never received any Fresh Start intervention for various reasons	329	Never received
Fresh Start not required	Never received any Fresh Start intervention because they were reading at the correct level when starting in year 5	159	Never received
Received phonics instead	Were included in Phonics classes due to either ability or staff shortage	36	Never received
Received Fresh Start	Received Fresh Start intervention	330	Received
Fresh Start teaching suspended	May have started receiving intervention, but have stopped teaching	15	Received
Unknown	Fresh Start UPNs not received	205	No data
Unknown – did not receive	Some Fresh Start UPNs received – those who did not receive no longer on system	137	Never received

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

The Education Endowment Foundation

5th Floor, Millbank Tower

21–24 Millbank

London

https://educationendowmentfoundation.org.uk

@EducEndowFoundn

Facebook.com/EducEndowFoundn