Peer Assisted Learning Strategies for Reading UK (PALS-UK)
Evaluation Protocol

Evaluator (institution): RAND Europe
Principal investigator(s): Sashka Dimova

PROJECT TITLE	Peer Assisted Learning Strategies for Reading UK (PALS-UK)
DEVELOPER (INSTITUTION)	Nottingham Trent University, Coventry University and Vanderbilt University
EVALUATOR (INSTITUTION)	RAND Europe
PRINCIPAL INVESTIGATOR(S)	Dr Alex Sutherland[1] (January 2019 – June 2019) Dr Ilie Sonia[2] (June 2019- July 2020) Dr Sashka Dimova (July 2020 – present)
PROTOCOL AUTHOR(S)	Dr Sashka Dimova Dr Alex Sutherland Dr Andreas Culora
TRIAL DESIGN	Two -arm cluster randomised controlled trial with random allocation at school level
PUPIL AGE RANGE AND KEY STAGE	Whole-class intervention, focus of the evaluation on Year 5 pupils (9-10 years old)
NUMBER OF SCHOOLS	89
NUMBER OF PUPILS	~2500
PRIMARY OUTCOME	Pupil reading attainment in Year 5
SECONDARY OUTCOME(S)	Oral reading fluency; Reading comprehension; and Self-efficacy in reading.

Protocol version history

VERSION	DATE	REASON FOR REVISION
1.0 [*original*]		*[leave blank for the original version]*
1.1	March 2021	*To capture changes to trial design and project timelines in response to the COVID-19 pandemic*

[1] Formerly RAND Europe, currently Behavioural Insights Team.
[2] Formerly RAND Europe, currently University of Cambridge.

List of abbreviations

DREC	Departmental Research Ethics Committee
EAL	English as Additional Language
EEF	Education Endowment Foundation
FIML	Full Information Maximum Likelihood
FSM	Free School Meals
GDPR	General Data Protection Regulation
HSPC	Human Subject Protection Committee
ICO	Information Commissioner's Office
IDEA workshop	Intervention Delivery and Evaluation Analysis Workshop
IPE	Implementation and Process Evaluation
ISCRTN	International Standard Randomised Controlled Trial Number
MAR	Missing at Random
MDES	Minimum Detectible Effect Size
MI	Multiple Imputation
NPD	National Pupil Database
PALS-UK	Peer Assisted Learning Strategies for Reading UK
PiRA	Progress in Reading Assessment
RA	Research Assistant
RS	Rising Stars
SEND	Special Educational Needs and Disability
TA	Teaching Assistant
TIDieR	Template for Intervention Description and Replication
TOC	Theory of Change
WIAT-III	Wechsler Individual Achievement Test - III
WWC	What Works Clearinghouse

Table of contents

Intervention

Peer Assisted Learning Strategies for Reading UK (PALS-UK)

Peer Assisted Learning Strategies for Reading UK (PALS-UK) is a whole-class, structured paired reading intervention that is designed to improve children's reading fluency and reading comprehension skills. The programme is aimed at children in Year 5. Pupils are first trained in the PALS-UK approach, paired up by their class teacher and then take it in turns to act as reader and coach in a set of structured activities. PALS-UK sessions comprise of four activities: (i) *partner reading*, in which the coach corrects the reader's reading errors using the PALS check-it procedure; (ii) *re-tell*, in which the reader re-tells a story with prompts from the coach; (iii) *paragraph shrinking*, in which both pupils summarise what they have read using up to ten words; and (iv) *prediction relay*, in which both pupils have to predict what will take place in the next half-page of a story after reading each half page. As an intervention, PALS-UK lasts 20 weeks, with sessions taking place three times a week, each for 30-35 minutes.

The Education Endowment Foundation (EEF) has granted funding to RAND Europe (the Evaluation Team) and Nottingham Trent University and Coventry University (the Delivery Team) to conduct a cluster-randomised controlled trial in primary schools in England assessing the effectiveness of PALS-UK.

In this evaluation, PALS-UK has been delivered to children in Year 5 classes in approximately 80 schools in England. Approximately half of the schools are from the Midlands and half are in the North East of England. Schools have been recruited by the Delivery Team but have been allocated to either the intervention or control group by RAND Europe. It has been recommended that PALS-UK sessions replace guided reading or other supplementary literacy activities (such as whole class reading) that a school currently does, but this has been subject to school preference.

For the PALS-UK implementation, teachers attended initial training delivered by Prof Kristen McMaster (University of Minnesota), Dr Emma Vardy (Nottingham Trent University) and Dr Helen Breadmore (Coventry University). Research Assistants (RAs) were recruited and trained by the Delivery Team to monitor implementation and provide support and feedback to the teachers.

The implementation of PALS-UK involves the following activities:

- All Year 5 teachers in intervention schools receive the PALS-UK manual and resources (including a selection of 20 reading books for children) that are needed to deliver the intervention.
- All teachers from Year 5 intervention schools attend two face-to-face training sessions: one full day session and one half day top-up session. The full day initial training takes place prior to implementation, and will be delivered by Prof. Kristen McMaster (a PALS trainer from the United States), Dr Emma Vardy and Dr Helen Breadmore. The initial training includes a background about the skills that underpin reading comprehension, detailed description of the intervention, practising the activities, pairing pupils and the sharing of hints/tips to support implementation. There are five sessions of this training delivered in a range of locations and teachers could decide which they attend. Training is delivered to up to 40 teachers. The top-up training takes place after teachers have delivered the 4 weeks of whole class training. Top-up training focuses on book selection, how to change pairs, and sharing good practice and will be delivered by Dr

Emma Vardy, Nottingham Trent University and Dr Helen Breadmore, Coventry University.

- Children training lasts for four weeks, and is provided by their Year 5 teachers. Children learn how to deliver the activities and how to work well in pairs. This training is part of the PALS-UK sessions, with a new skill introduced during each session. Each lesson is fully manualised and all resources are provided in the PALS-UK manual.
- Children swap pairs every four weeks. More information on pupil pairing is provided below.
- Children work in pairs for 30-35 minutes, three times a week during the autumn, winter and spring terms of the school year 2019/2020. This will continue for 16 weeks after the initial training. Two of these sessions will be observed by an RA and two will be observed by a peer teacher. Peer teachers (i.e. another Year 5 teacher or member of staff in the same school) and RAs observe four of the classroom sessions.[3] During the visit they assess whether the PALS-UK procedures and activities are implemented correctly. More information on the content and timing of the observations is provided in *Review of RA and peer classroom observation* on p.25.

Pupils are arranged by their teachers into pairs, and teachers should change pairings every four weeks. Students are paired based on ability. The teacher orders the pupils based on ability and divides the group by the median of the class, based on their judgement. Then, the strongest reader of the above-median group is paired with the strongest reader of the below-median group. Thus, there is some gap in reading ability between the pupils, but not too large. Teachers are able to reorganise pupils depending on their judgement about whether they expect the two to work well together. Pupil pairings is displayed on the board so pupils know who it is that they are paired with. Teachers record the pairings in the classroom. Teachers select the reading material for each pair, based on the children's reading ability. Pupils with severe Special Educational Needs and Disability (SEND) who require the support of a full-time teaching assistant (TA) can be paired with their TA. Furthermore, teachers are advised to re-pair students if something does not work well. Instructions on how to pair children are included in the manual. At the top-up training, one of the activities will be to re-pair their class, so that teachers are supported in this activity.

The logic model (see

[3] In single form entry schools, the peer teacher may instead be a Year 4 or Year 6 teacher or literacy lead.

below) describes the linkages between the programme activities and expected outcomes. Teaching children to read fluently and comprehend texts is fundamental for school achievement.[4] Reading fluency strategies are used to increase overall reading confidence and comprehension.[5] The repeated opportunity in PALS-UK to practice oral reading and receive feedback with accuracy checking is expected to train oral reading fluency. This is all three elements of oral reading fluency; accuracy, automaticity and prosody. Additionally, PALS-UK is intended to develop comprehension through three specific activities (re-tell, paragraph shrinking and prediction relay), which are expected to train comprehension, automaticity and prosody. It is expected that increasing the practice of reading aloud and completing reading comprehension activities with structured peer support will help pupils' self-efficacy. Being both the coach and reader is also expected to develop confidence levels for both the weaker and stronger reader. In theory, PALS-UK improves reading fluency and comprehension, which influences both self-efficacy and peer relationships. In the long term, fluency and comprehension are expected to influence overall reading attainment.

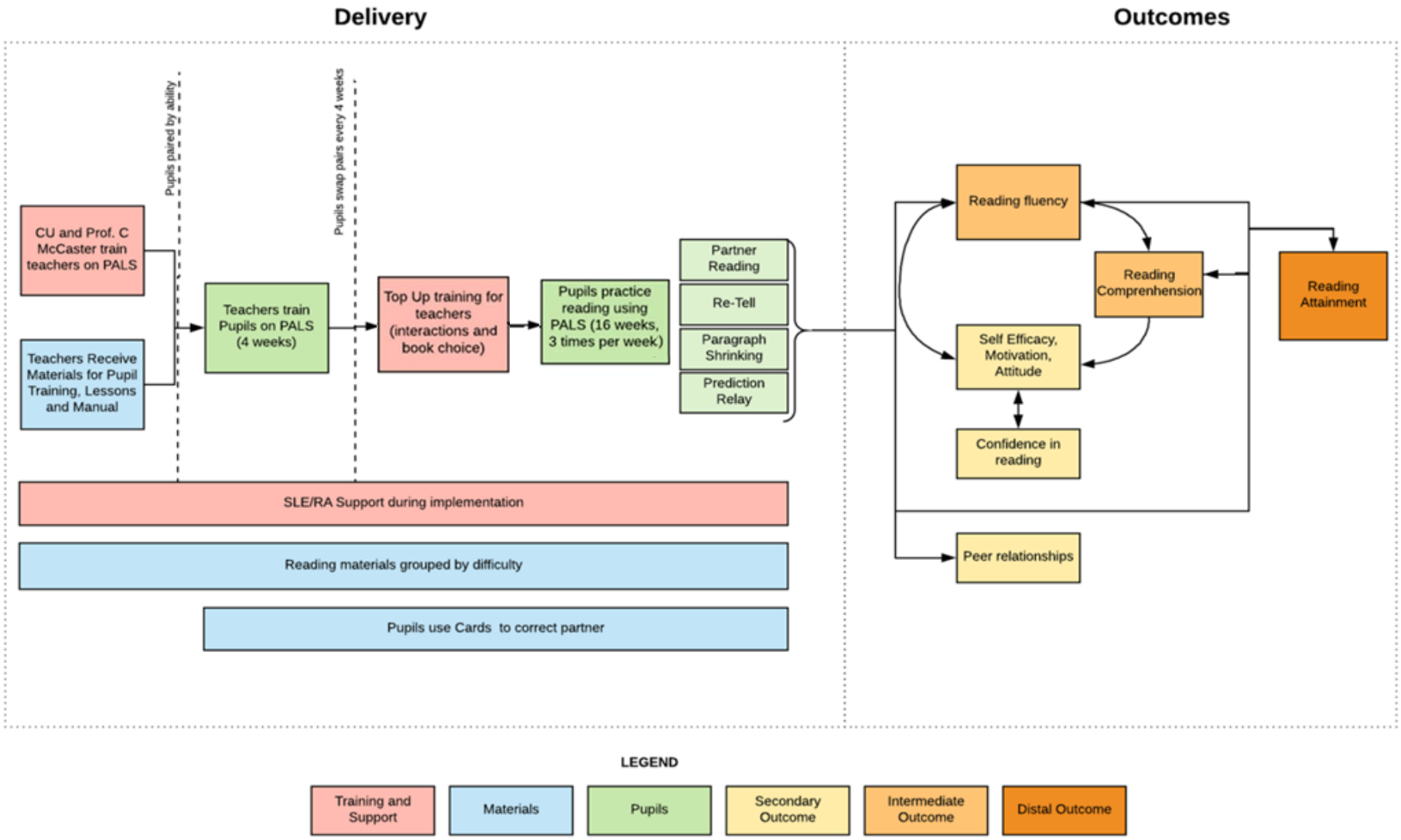
To enable effective delivery, Year 5 teachers have been provided with ongoing support from Nottingham Trent University and Coventry University. Other strategies which were in place to ensure high fidelity include the following:

- Videos of how to deliver the PALS-UK activities were shared with teachers
- E-mails to teachers with reminders to switch pairs and tips for book selection
- Just-in-time support offered by RAs to teachers, where teachers could get in touch with an RA via email for additional specific support

The intervention has been delivered using resources developed by Nottingham Trent University/Coventry University with the support of programme developers from the United States. The Grade 2-6 American version of PALS is available to purchase from Vanderbilt University for the cost of £31. There are other programme costs of approximately £150 per teacher that are related to teacher training and provision of resources. The total costs for the intervention are estimated at approximately £200. Following a meeting in November 2020, it was agreed to waive any fee to the intervention schools taking part in the delayed outcome testing. Control schools will receive a payment of £500 per school upon completion of outcome testing in the trial. .

[4] Hulme, C., and Snowling, M. J. (2011). Children's reading comprehension difficulties: nature, causes, and treatments. Curr. Direct. Psychol. Sci. 20, 139–142. doi: 10.1177/0963721411408673
[5] Rasinski, T. V. (2003). *The fluent reader: Oral reading strategies for building word recognition, fluency, and comprehension*. Scholastic Inc.

Figure 1. PALS Logic Model



Study rationale and background

PALS is a programme originally developed in the United States by Douglas Fuchs, Lynn Fuchs and other researchers from Vanderbilt University, with the main aim to improve reading fluency and comprehension in the Grade 2-6 version.[6][7] This trial is important because peer tutoring interventions are appealing due to their potential impact and low cost. Academic research and the EEF Teaching and Evidence Toolkit reports evidence of moderate/high effects for peer tutoring, including interventions such as PALS, as well as cross-age tutoring.[8,9] PALS has been widely implemented and studied in the US but a What Works Clearinghouse (WWC) effectiveness summary for 'beginning readers' found only three studies met WWC evidence standards.[10] The summary suggested that between 13 and 20 weeks of PALS had mixed effects on comprehension and no discernible effect on fluency for pupils in kindergarten and first grade. A WWC summary on the topic of PALS for older children included only one study, focusing on nine and ten-year old pupils who received 15 weeks of PALS that met their evidence standards.[11] As such, rigorous evidence on the effectiveness of PALS is still needed.

Although there is a long history of peer tutoring interventions in the UK, results from recent empirical studies have not been encouraging.[12] This reinforces the need to understand how the design of interventions influences their results, which must involve a stronger emphasis on theorising and testing causal mechanisms in relation to peer tutoring interventions.[13,14] In the UK, the evidence is stronger for *cross*-age than *same*-age tutoring and there is little UK research that specifically evaluates the PALS intervention.[15] Accordingly, PALS-UK is a strong candidate for evaluation as it provides structure for within-class, *same*-age peer interactions, which makes scale-up more feasible.

PALS-UK would also benefit from an independent evaluation as previous evaluations of PALS have been developer-led, meaning their results may be atypical because of the greater attention paid to implementation that close monitoring by developers can lead to. Additionally, PALS-UK is one of the few practices that focus on fluency as a key component in reading and

---

[6] https://vkc.mc.vanderbilt.edu/frg/what-is-pals/

[7] Reading fluency is multifaceted concept: it is the speed, accuracy and prosody(expression) used when reading a text. For more information see Kuhn, M. R., Schwanenflugel, P. J., & Meisinger, E. B. (2010). Aligning theory and assessment of reading fluency: Automaticity, prosody, and definitions of fluency. *Reading Research Quarterly*, *45*(2), 230-251.

[8] Education Endowment Foundation (2018). *The EEF Teaching and Evidence Toolkit.* London: Education Endowment Foundation. Retrieved from: https://educationendowmentfoundation.org.uk/evidence-summaries/teaching-learning-toolkit/peer-tutoring/

[9] Additional academic references can be found in: Topping, K., Millder, D., Thurston, A., McGavock, K., & Conlin, N. (2011). Peer tutoring in reading in Scotland: thinking big. Literacy, 45(1), 3–9; Lloyd, C., Edovald, T., Kiss, Z., Morris, S., Skipp, A., & Ahmed, H. (2015). Paired Reading: Evaluation report and Executive Summary. London: Education Endowment Foundation. Retrieved from: https://v1.educationendowmentfoundation.org.uk/uploads/pdf/Paired_Reading.pdf;
Lloyd, C., Edovald, T., Morris, S., Kiss, Z., Skipp, A., Haywood S. (2015b). Durham Shared Maths Project: Evaluation report and Executive Summary. London: Education Endowment Foundation. Retrieved from: https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/Shared_ Maths.pdf;
Shenderovich, Y., Thurston, A., & Miller, S. (2016). Cross-age tutoring in kindergarten and elementary school settings: A systematic review and meta-analysis. International Journal of Educational Research, 76, 190-210.

[10] WWC. (2012a). Peer-Assisted Learning Strategies. Available from: https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_pals_060512.pdf

[11] WWC. (2012b). Peer-assisted learning/literacy strategies. Available from: https://files.eric.ed.gov/fulltext/ED531596.pdf

[12] Lloyd, C., Edovald, T., Kiss, Z., Morris, S., Skipp, A., & Ahmed, H. (2015). Paired Reading: Evaluation Report and Executive Summary. *Education Endowment Foundation*.

[13] Morris, S. P., Edovald, T., Lloyd, C., & Kiss, Z. (2016). The importance of specifying and studying causal mechanisms in school-based randomised controlled trials: lessons from two studies of cross-age peer tutoring. *Educational Research and Evaluation, 22*, 422-439.

[14] Lortie-Forgues, H., & Inglis, M. (2019). Rigorous Large-Scale Educational RCTs Are Often Uninformative: Should We Be Concerned?. *Educational Researcher*, 0013189X19832850.

[15] Tymms, P. Merrell, C., Thurstone A., Andor, J., Topping, K., & Miller, D. (2011). Improving attainment across a whole district: school reform through peer tutoring in a randomized controlled trial. *School Effectiveness and School Improvement*, *22*(3), 265-289.

in learning to read. It can therefore provide evidence on the nature of effective instruction in fluency.[16] This EEF-funded efficacy trial will help determine whether PALS-UK leads to observable improved outcomes for pupils.[17]

Impact Evaluation

Research questions

This project seeks to answer the following research questions:

Does the implementation of PALS-UK result in:

1. Higher levels of reading attainment by pupils in treatment schools compared to pupils in control schools?
2. Higher levels of fluency and oral skills for pupils in treatment schools compared to pupils in control schools?
3. Higher self-efficacy in reading for pupils in treatment schools compared to pupils in control schools?
4. Higher outcomes in all above areas for certain groups of pupils (including: pupils eligible for Free School Meals; pupils for whom English is another language (EAL), and pupils with lower or higher reading ability) in treatment schools compared to the same groups of pupils in control schools?

The impact evaluation is designed to explore outcomes for schools that implement PALS compared to schools running business as usual.

Owing to changes to the design outlined below (see COVID-19 driven changes) the impact evaluation research questions were modified. The research hypotheses this trial is able to address are as follows:

Hypothesis 1 (Primary Outcome): Year 5 pupils in randomly allocated primary schools participating in PALS-UK (intervention schools) will have higher levels of reading attainment than pupils in control schools by Autumn term 2020, as measured by Progress in Reading Assessment (PiRA).

Hypothesis 2 (Secondary outcome): Year 5 pupils in randomly allocated primary schools participating in PALS-UK (intervention schools) will have higher self-efficacy in reading than pupils in the control schools by Autumn term 2020, as measured by pupil self-efficacy in reading.

Hypothesis 3: Year 5 pupils in treatment schools for whom English is another language (EAL) will perform better on all outcomes that EAL pupils in control schools**.**

Hypothesis 4: Year 5 pupils in treatment schools who are entitled to Free School Meals will perform better on all outcomes than FSM pupils in control schools.

[16] Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific studies of reading*, *5*(3), 239-256.
[17] Flay, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S., ... & Ji, P. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. Prevention science, 6(3), 151-175.

Hypothesis 5: Lower and higher ability Year 5 pupils in treatment schools will perform better on all outcomes than pupils of similar ability in control schools.

COVID-19 driven changes

Testing timeframes

Due to the COVID-19 pandemic, planned outcome testing in the early Summer term 2020 was postponed until December 2020. The outcome measures remained the same. Testing was organised in as much as possible to match the original plans (i.e. whole-class testing for the primary outcome measure and the reading self-efficacy secondary outcome; one-to-one testing with a sample of ten randomly selected pupils per school for the oral fluency and reading comprehension secondary outcome based on the WIAT III). The tests remained the same to account for the disruption caused by COVID-19, despite the fact that children were older than originally planned at test administration.

Test administration

Due to the schools' no visitor policy on site unless necessary, the PiRA test was administered by teachers, but was marked by skilled markers who were blinded about the allocation. The WIAT III was administered virtually by qualified CL2R[18] professionals.

Changed to secondary outcome impact evaluation

Testing for the primary outcome for those schools that were willing to participate in delayed outcome testing was completed in December 2020. However, due to the movement of schools to remote learning as a result of the national lockdown from January 2021, completion of one-to-one testing for the Oral reading fluency and Reading comprehension subtests of the WIAT-III was no longer considered feasible. Owing to the minimal power and the likelihood of higher attrition following lockdown, the decision was taken to stop all data collection activities for assessing oral fluency as measured by the WIAT-III subtests. Thus, while data was collected in a small number of schools (n=22) on the WIAT- III reading comprehension subtest before school closures were implemented, collection of data on the secondary outcome on oral fluency (i.e. WIAT III) was not completed. The secondary outcome on self-efficacy was collected in majority of schools in December 2020, and will be analysed as originally planned.

Details for the complete secondary analyses that had been planned can be found in the original protocol.

Design

Trial type and number of arms	Two-arm, stratified and cluster-randomised trial at the school level
Unit of randomisation	Schools

[18] CL2R qualification is qualified teaching status and a further Post Graduate Qualification in SEN i.e Post Graduate Diploma or Masters

Stratification variables (if applicable)		Region (Midlands and North East) School size (single-form entry versus multiple-form entry)
Primary outcome	variable	Reading attainment
	measure (instrument, scale)	Progress in Reading Assessment (PiRA)[19]
Secondary outcome(s)	variable(s)	Oral reading fluency; Reading comprehension; and Self-efficacy in reading.
	measure(s) (instrument, scale)	WIAT-III: Oral reading fluency and reading comprehension subtest Self-efficacy survey[20]

UPDATE: Given school closures in January 2021 and extended lockdown due to the COVID-19 outbreak, it was not possible to continue collecting secondary outcome data on oral reading fluency and reading comprehension using the WIAT-III. Data was collected in a small number of schools (n=22) on the WIAT-III reading comprehension subtest before school closures were implemented. Thus, the WIAT-III data collection listed above was not completed.

The PALS-UK evaluation has been designed as a two-group parallel, stratified, cluster-randomised trial, with school as the unit of randomisation[21]. To ensure comparability of schools in the intervention arm and the control arm, schools have been randomised within regions and by school size (single-form vs multi-form entry) in order to balance the study arms on geographical location and the overall size of the school.[22]

Multi-entry schools

Intervention schools: If the school has been selected to deliver PALS-UK, all Year 5 teachers at the school will receive the PALS-UK manual, training, resources and support needed to deliver the intervention. Intervention schools with two or multiple-entry Year 5 classes will deliver the intervention in all classes, but only one randomly selected Year 5 class will be tested.

Control schools: Multi-entry control schools will continue implementing business as usual reading practices and only one randomly selected Year 5 classroom will undertake baseline and post-trial outcome testing. This significantly reduces testing costs but has only a small effect on the power of the study.

Classes that need to complete testing have been randomly selected before programme delivery and before school randomisation. The same selected Year 5 class will be asked to complete a reading attainment test at the beginning and end of the project. Schools in the

[19] https://www.risingstars-uk.com/series/assessment/rising-stars-pira-tests
[20] Carroll, J. M., & Fox, A. C. (2017). Reading self-efficacy predicts word reading but not comprehension in both girls and boys. *Frontiers in psychology*, *7*, 2056.
[21] More information on randomisation is available at the study Statistical Analysis Plan: https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/pals-uk/

[22] Oakes, J. M. (2013). Effect identification in comparative effectiveness research. *eGEMs*, *1*(1).

control group will receive a payment (£500) for participating in the trial after outcome testing has been completed (as an incentive to remain in the trial).

Selection of sub-sample of children

Around one third of children in the classroom have been randomly selected to undertake an individually administered oral fluency subtest (see below Fluency and language Comprehension). In multi-entry schools the sub-sample of children have been selected within the classroom selected for baseline outcome testing. The sub-sample of Year 5 children have been selected at the beginning of the trial by RAND Europe. However, the sub-sample of selected children who will complete the individually administered comprehension and fluency test will be revealed only at the time of the post-trial outcome testing. This minimises the possibility that schools may inadvertently focus resources/effort on the children who will complete the individual fluency and comprehension test if that is known too far in advance.

Randomisation

Randomisation has been conducted in Stata by the evaluation team. Randomisation has been stratified on region and school size. Stratification by region was necessary as half of the schools will be recruited from the Midlands and another half from the North East. In addition, school size was felt to be an important factor in terms of both delivery and outcomes (operationalised as single-entry versus multiple-entry school).

Randomisation occurred at the start of the school year 2019/2020, with allocation revealed to schools on a school-by-school basis once each has completed their baseline testing of pupils. The trial allocation was recorded and communicated to the Delivery Team after the baseline testing was completed.

Baseline equivalence will be examined based on the initial randomisation. A well-conducted randomisation will yield groups that are equivalent at baseline.[23] Since schools will be randomly allocated to the control and intervention conditions, any imbalance at baseline will have occurred by chance. To assess imbalance at baseline, we will compare groups at school and pupil levels, by means of cross-tabulations and histograms that assess the distribution of each characteristic within the control and treatment groups.

Participants

Schools have been recruited by the delivery team, based on the following eligibility criteria:

- Schools have not used PALS in the last three years.
- Schools have not taken part or will not take part in other EEF trials for KS2 in the school year of 2019/2020 and the first half of the school year 2020/2021.
- Schools should not have mixed age classes unless agreed between the delivery and evaluation team to include them [24].There must be at least a one single-form entry Year 5 classroom in the school.

[23] Glennerster, R., & Takavarasha, K. (2013). *Running randomised evaluations: A practical guide.* Princeton: University Press.

[24] Exception is made for classes where the mixed years are taught separately for English and Maths so in effect they are one age class. For example, Year 5 and Year 6 pupils are mixed but in the morning Year 5 are taught separately from the Year 6.

- Schools are willing to provide child background information to the delivery team (as specified in the Memorandum of Understanding).
- Schools will support testing two times within the project's timeline (tests to be administered in intervention and control schools).
- Schools are willing to engage with the intervention and to implement the intervention.
- Schools are willing to facilitate data collection by the evaluation team.

Participating schools have signed a Memorandum of Understanding (MoU), which outlines the roles and responsibilities of all stakeholders involved. The MoU makes it clear that once schools agree to participate, the expectation will be that final outcome testing of pupils will take place even if the school withdraws from the intervention. Furthermore, schools have been asked to consent to sharing the specified data with the delivery team. Schools will be asked to notify the delivery and/or research team immediately if a teacher or pupil is withdrawn from data collection.

The MoU and the parent information sheet have been appropriately updated and school staff and parents have been informed about changes to testing and data collection due to the COVID-19 pandemic.

Teachers

All Year 5 teachers from intervention schools were trained, but outcomes are measured for one randomly selected class only. Year 5 teachers randomised into the control group did not receive any training and instead have been taking a "business as usual" approach to teaching.

Children

There are no inclusion/exclusion criteria based on pupil characteristics as PALS-UK is a whole-class intervention. To minimise burden on schools, pupils enrolled at the time of school recruitment in 2019 are included in the study, but pupils who join the schools at a later time will not be included in the evaluation as this would require additional information collected from schools. This will be checked by comparing pupil names/UPN at the start and end of the study to ensure that only those pupils present from the start will be included. Additionally, children with severe Special Educational Needs and Disabilities (SEND) have not been necessarily paired with other children but could carry out PALS-UK activities with a Teaching Assistant (TA).

If parents choose to withdraw their child from the study, their data will not be collected. However, as this is a whole class intervention, children still took part in the PALS-UK session but they have not be assessed. Parents have the right to withdraw their child at any time throughout the study duration.

Sample size calculations

		OVERALL	FSM
MDES		0.228	0.263
Pre-test/ post-test correlations	level 1 (pupil)	0.7	0.7
	level 2 (class)	NA	NA

	level 3 (school)	0	0
Intracluster correlations (ICCs)	level 2 (class)	NA	NA
	level 3 (school)	0.13	0.13
Alpha		0.05	0.05
Power		0.8	0.8
One-sided or two-sided?		Two	Two
Average cluster size		28	7
Number of schools	Intervention	45	45
	Control	45	45
	Total	90	90
Number of pupils	Intervention	1260	315
	Control	1260	315
	Total	2520	630

We have based the power calculations on both the information provided in the Invitation to Tender and the subsequent set-up meetings with the Delivery team (Coventry University) and the EEF. Power and minimum detectable effect size (MDES) calculations were performed using PowerUp.[25]

The power calculations assume equal allocation to treatment and control groups. The estimation at the protocol stage is based on 90 schools [26]. To reduce the testing burden and overall study costs, RAND Europe randomly selected one class in schools with multiple entry Year 5 classes to complete outcome testing. With one Year 5 class per school included in the evaluation, we have assumed an average cluster size of 28 pupils, which should give us a buffer for attrition at pupil level given that average class size is 30. Based on previous research, we assume the proportion of variance in Level 1 outcomes explained by Level 1 covariates R_1^2 of 0.49 (equating to a correlation of 0.70, as per the power calculation table above) and R_2^2 of 0.00.[27] We use two-level clustered designs and base our calculations on intra-cluster correlation ICC of 13 per cent. Assuming a desired power of 80 per cent, alpha of 5 per cent, an ICC of 0.13 and a continuous, normally distributed outcome, the MDES is *d*=0.228. Using the parameters above and assuming that on average there are around seven FSM pupils, the MDES is 0.263. As such, even though considered an efficacy trial, the study should be powered to detect meaningful differences between groups.

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

[26] Until now 98 schools were recruited in the trial but estimations are assuming 90 schools in the study. The conservative assumptions build in a buffer for attrition at school and pupil level. Power calculations on actual sample of recruited schools and actual average class size will be presented in the statistical analysis plan.

[27] Choudry, S., Squires, G., and Humphrey, N. (2017). *Statistical Analysis Plan for Achievement for All*. Retrieved from: https://educationendowmentfoundation.org.uk/public/files/Projects/Round_9_-_Achievement_for_All_SAP_(amended).pdf

Due to the COVID-19 pandemic, we expect that there may be a higher level of attrition at the post-testing stage. The updated Statistical Analysis Plan of the study presents the potential impact of different levels of attrition on the MDES.

Outcome measures

Primary outcome measure: reading skills

The primary outcome will be a standardised measure of reading skills of pupils based on the Progress in Reading Assessment (PiRA) test (Hodder Education), which is sold in the UK by Rising Stars (RS) Assessment (provider for primary schools). The test is advertised as taking around 40 minutes to complete.[28] This test evaluates: *general reading skills* (specifically comprehension, inference, and language, structure and presentation). The primary outcome measure will be administered at two different time points. Baseline testing will take place in the first two weeks of September 2019 (2nd to the 13th September 2019), while post-trial outcome testing will take place in the Autumn term 2020.

Secondary outcome measure

1. Fluency and language Comprehension

The secondary outcome of oral fluency and language comprehension will be based on the oral fluency and reading comprehension subtests of the Wechsler Individual Achievement Test–III (WIAT-III).[29,30] The oral fluency subtest measures: speed, accuracy, prosody, and fluency in reading, while the reading comprehension scale measures: comprehension of various types of text, including stories, advertisements and how-to passages.[31] These subtests ask the pupil to read a passage aloud (or silently in the reading comprehension subtest) and then orally respond to comprehension questions. The subtests are administered individually at the end of the trial in Autumn term 2020 and will take approximately 30 minutes for each pupil to complete.

To reduce the testing burden and costs, it was agreed to administer the WIAT-III subtests to ten randomly selected children. Pupils have been randomly sampled at the beginning of the trial. These children were selected within the randomly selected class who will complete both baseline and post-trial outcome testing. More information on the selection of the sub-sample is presented in *Selection of sub-sample of children* on p. 13.

Update: In December 2020, schools started completing elements of the WIAT-III with ten randomly selected children. In total, 22 schools completed the WIAT III reading comprehension subtest. The test was administered online. However, due to the movement to remote learning in January 2021 as a result of national lockdown it was not feasible to proceed with testing. Given the probability of high attrition and the minimal power associated with the

[28] https://www.risingstars-uk.com/series/assessment/rising-stars-pira-tests
[29] https://www.pearsonclinical.co.uk/Sitedownloads/Productpdfs/wiat-iii-uk-subtests-description.pdf
[30] Burns, T. G. (2010). Wechsler Individual Achievement Test-III: What is the 'Gold Standard' for Measuring Academic Achievement?. *Applied Neuropsychology*, *17*(3), 234-236.
[31] Prosody refers to reading expression; the ability to incorporate the rise and fall of pitch, to pause within and between sentences and etc. For more information look at: Schwanenflugel, P. J., Hamilton, A. M., Kuhn, M. R., Wisenbaker, J. M., & Stahl, S. A. (2004). Becoming a fluent reader: reading skill and prosodic features in the oral reading of young readers. *Journal of educational psychology*, *96*(1), 119.

given sample size (the test was administered to sub-sample of 10 children per school) the decision was taken to not collect impact outcomes on oral fluency (i.e. WIAT-III).

2. Self-efficacy

The evaluation will aim to understand if PALS-UK positively affects pupils' reading self-efficacy. Self-efficacy refers to children's beliefs in their capability to produce satisfactory attainment.[32] A self-report survey containing 20 items, developed by researchers from Coventry University, will be used to assess children's self-efficacy. The survey has been tested in a small study with thirty children and has an internal reliability of (α = 0.912). More information on the questionnaire is available in Carroll and Fox.[33]

To assess changes in pupil's self-efficacy the survey will be administered at the same time as the primary outcome assessment. It will be completed at two different time points (September 2019 and Autumn 2020). The survey will be compiled in machine-readable forms, to allow scanning, data entry and scoring by RAND Europe.

Testing time

The overall testing burden on pupils will be up to 50 minutes, while overall testing time for those 10 pupils selected for individual testing will be up to 80 minutes. That will be managed by ensuring that tests are organised either side of break-time, but that the primary outcome measure is completed first in all schools.

Test administration

All tests will be administered and scored by testers who have been trained by NTU and Coventry University and who will be blind to allocation. Baseline testing will be administered by classroom teachers while post-tests will be undertaken by testers selected by RAND.[34] RAND Europe's testers will be adequately trained to ensure consistency with baseline testing procedures. To ensure a minimum of 90 per cent completion rate and limit the amount of missing data, RAND Europe's testers will conduct two rounds of testing sessions: first to test all pupils; second to test pupils missed in the first round. At baseline, teachers did one more round of testing on pupils that have missed the initial baseline assessment. Baseline testing was completed before randomisation.

Optional follow-up

Long-term follow-ups of educational interventions are rare; as such this trial could also offer a good opportunity to assess the effectiveness of PALS-UK measured one year after the end of the intervention, once pupils have transitioned into Year 6. This opportunity – likely using KS2 outcome data – is currently being considered by the EEF and thus not discussed further in this protocol.

[32] Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary educational psychology*, *25*(1), 82-91.
[33] Carroll, J. M., & Fox, A. C. (2017). Reading self-efficacy predicts word reading but not comprehension in both girls and boys. *Frontiers in psychology*, *7*, 2056.
[34] It was felt that there is no need to provide additional training on baseline testing as teachers frequently administer PIRA tests in schools.

Analysis plan

Primary intention-to-treat (ITT) analysis

The outcome analysis will be on an intention-to-treat (ITT) basis.[35] The analysis will include all randomised schools/pupils in the groups to which they were randomly assigned, regardless of the treatment actually received, withdrawal from the intervention post-randomisation, or deviations in programme implementation. This principle is key in ensuring an unbiased analysis of intervention effects. This approach compares outcome means for the treatment and comparison groups, and subjects are analysed according to their randomised group allocation. The ITT approach is inherently conservative as it captures the averaged effect of *offering* the intervention, regardless of whether the participants complied with assignment.

The primary outcome is pupil-level PiRA test scores standardised with a mean of zero and standard deviation of one. To estimate the impact on the primary outcome we will use a two-level multilevel model to account for clustering of data. Multilevel approaches assume that the schools in the study are a random sample of all schools and that the multilevel modelling framework can flexibly handle complex variation within/between schools.[36,37,38]

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

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

where Y_{ij} is the PiRA score for child *i* in school *j*; β_0 is the cluster level coefficient for the slope of a predictor on language; PALS_j is a binary indicator of the school assignment to intervention [1] or control [0]; Z_j are school-level characteristics, here the two stratifying variables of geographical location and number of classes per schools (as used for randomisation); X_{ij} represents characteristics at pupil level (pupil $_i$ in school $_j$), specifically the pre-intervention PiRA score[39]; u_j are school-level residuals ($u_j \sim i.i.d\ N(0, \sigma_u^2)$) and e_{ij} are individual-level residuals ($e_{ij} \sim i.i.d\ N(0, \sigma_e^2)$). Equation (1) is known as a 'random intercepts' model because $\beta_{0\mathrm{j}} = \beta_0 + u_j$ is interpreted as the school-specific intercept for school j and $\beta_{0j} \sim i.i.d\ N(\beta_0, \sigma_u^2)$ is random as in, it is a number that can take any value).

Our target parameter (i.e. the focal result of the trial) is τ, a binary treatment/control indicator variable. That will tell us the average effect of the intervention on pupil outcomes in treatment schools compared to those in control schools.

---

[35] Fisher, L. D., Dixon, D. O., Herson, J., Frankowski, R. K., Hearron, M. S.,et al. (1990). Intention to treat in clinical trials. In K. E. Peace (ed). *Statistical Issues in Drug Research and Development* New York: Marcel Dekker, pp. 331–350.

[36] Hox, J. (1998). Multilevel modeling: When and why. In Balderjahn I., Mathar R., Schader M. (eds) *Classification, Data Analysis, and Data Highways. Studies in Classification, Data Analysis, and Knowledge Organization.* Berlin, Heidelberg: Springer.

[37] Snijders, T. A. (2005). Power and sample size in multilevel modeling. *Encyclopedia of statistics in behavioral science*, *3*, 1570-1573.

[38] Snijders, T. A., & Bosker, R. J. (1994). Modeled variance in two-level models. *Sociological methods & research*, *22* (3), 342-363.

[39] It would be included in a class-level average of the baseline in the final analysis

Secondary outcome analysis

The secondary outcomes will be based on the WIAT-III subtests and the self-efficacy survey. Pupil's oral fluency and reading comprehension will be assessed following the same specification to equation (1) listed under primary outcome analysis above, but we will substitute these measures as outcomes.

UPDATE: The secondary analyses based on the WIAT-III subtests were planned but are no longer possible as outcomes were collected post intervention in 22 schools for a single subtest of the WIAT III: Reading comprehension. However, we feel it will be beneficial to report mean outcomes on the Reading comprehension WIAT III subtest, despite the fact that the secondary analysis based on the WIAT-III that was originally planned is no longer possible.

Subgroup analyses

The study is not powered to detect significant differences between sub-groups. However, we will report mean outcomes by sub-categories of children who speak English as an additional language (EAL), children eligible for Free School Meals (FSM) and high vs. low reading achievers at baseline as a basic descriptive step.

The focus on EAL is important because of the low educational attainment of EAL and their increasing presence in the state school system.[40] Additionally, an earlier evaluation of a PALS programme in 10 schools in the US focused on English Language Learners (ELL) pupils with Learning Difficulties (LD) found a large positive effect.[41]

Findings about the effects of pairing low with high ability pupils are mixed.[42] The evaluation of PALS-UK will inform if there are any differences in programme effect for lower ability pupils compared to higher ability pupils. Pupil ability will be based on the baseline data on the PiRA reading assessment. We will consult with the test developers before we define a threshold for low and high achievers.

As an exploratory modelling approach, EAL will be incorporated into the regression analysis as a binary variable [1] if a child is considered EAL and as [0] if they are not considered EAL. The EAL indicator will then be interacted with treatment allocation to assess the conditional impact of PALS on EAL pupils. We will follow the same strategy for FSM pupils [yes/no] (using the EverFSM indicator as the FSM variable).

Finally, we will explore differential effects for FSM pupils as they are considered a key target group by the EEF. We will analyse FSM pupils as a separate subsample. We will estimate programme impact on the primary outcome for FSM pupils using a separate model.

As these analyses are exploratory and underpowered, we will report point estimates and confidence intervals but will not report significance tests/p-values.

Additional analyses

Imbalance at baseline

We will take an active approach to address imbalance by stratifying the randomisation. Balance checks will be conducted at the school and pupil level.

[40] Strand, S., Malmberg, L., & Hall, J. (2015). English as an Additional Language (EAL) and educational achievement in England: An analysis of the National Pupil Database.
[41] Sáenz, L. M., Fuchs, L. S., & Fuchs, D. (2005). Peer-assisted learning strategies for English language learners with learning disabilities. *Exceptional children*, *71*(3), 231-247.
[42] Lou, Y. Abrami, P. C. Spence, J. C. Poulsen, C. Chambers, B. d' Apolonia, S. Within-class grouping: A meta-analysis Review of Educational Research 66 423–458 1996.

At the school level, we will check the balance in the following variables by means of cross-tabulations and histograms that assess the distribution of each characteristic within the control and intervention groups aggregated from the data in the study sample (rather than from publically available school-level statistics):

- Proportion of children in Year 5 class speaking English as an additional language (EAL).
- Average baseline PiRA reading score of pupils in Year 5 class.
- Proportion of children in focal Year 5 class eligible for FSM.
- Proportion of all children in the school eligible for FSM.

At the individual (pupil) level, balance will be assessed for the following characteristics:

- EAL status.
- FSM status for pupils.
- Gender.
- Average age in months.
- PiRA baseline score.

Statistical significance tests will not be carried out to assess the balance, as their premise does not hold in randomised control trials.[43] Instead, tables of the means (and standard deviation, where appropriate) for each characteristic will be presented along with distributions.[44,45]

In order to establish baseline equivalence between the intervention and comparison groups we will estimate the effect size based on the average baseline PiRA reading score. It will be calculated as the mean difference in the PiRA score for Year 5 between pupils in primary schools participating in PALS-UK (intervention schools) and pupils in control schools, divided by the pooled standard deviation. We will consider the groups equivalent at baseline if the effect size based on the PiRA is 0.05 or less in absolute value.

Effect size

We will use the effect sizes for cluster-randomised trials given in the EEF evaluator guidance – an example, adapted from Hedges, is given below:

$$(2)\ ES = \frac{(\bar{Y}_T - \bar{Y}_C)_{adjusted}}{\sqrt{\sigma_S^2 + \sigma_{error}^2}}$$

Where $(\bar{Y}_T - \bar{Y}_C)_{adjusted}$ is the mean difference between intervention groups adjusted for baseline characteristics and $\sqrt{\sigma_S^2 + \sigma_{error}^2}$ is an estimate of the population standard deviation (variance). The effect size therefore represents the proportion of the population standard

[43] Baseline Data. *Consort 2010*. Retrieved 26 March 2019: http://www.consort-statement.org/checklists/view/32-consort/510-baseline-data

[44] There is a convention in some disciplines that a 10pp (or larger) difference in treatment and control means at baseline constitutes 'imbalance' is thus justification for including those measures in sensitivity analyses, but there are counter-arguments to this idea. See Roberts, C. and Torgerson, D. (1999) 'Baseline imbalance in randomised controlled trials', *BMJ*, 319:185; de Boer et al. (2015) 'Testing for baseline differences in randomized controlled trials: an unhealthy research behavior that is hard to eradicate', *International Journal of Behavioral Nutrition and Physical Activity*, 12:4.

[45] Senn, S. (1994) 'Testing for baseline balance in clinical trials', *Statistics in Medicine*, 13: 1715-1726.

deviation attributable to the intervention.[46] The exact effect size used will depend on whether there are equal or unequal sample sizes in trial arms.

The same approach will be used for primary and secondary outcomes.

Missing Data

If schools drop out of the intervention, we will still aim to complete outcome testing in all schools and that will form part of the MOU with schools. If pupils move schools we will not attempt outcome testing for those pupils.

We will explore attrition across trial arms as a basic step to assess bias.[47] To assess whether there are systematic differences between those who have contributed data and those who have not and whether factors should be included in analysis, we plan to model missingness at follow-up as a function of baseline covariates, including treatment. For item non-response, the extent of missingness may in part determine the analytical approach. For less than five per cent missingness overall, a complete-case analysis would suffice, regardless of the missingness mechanism.[48] Our default would be to check results using approaches that account for missingness which rely on the weaker Missing at Random (MAR) assumption, building the MAR conditioning variables from our initial work predicting missingness.

If there was systematic missingness of predictor variables, for example, we would explore options for using full information maximum likelihood (FIML) and/or multiple imputation (MI). [4950] In the event that the more detailed pupil baseline (pre-test) data are unavailable for individual pupils (or even a whole school), those pupils would be included in the outcome analysis via FIML, rather than sacrifice statistical power through excluding them.

Implementation and process evaluation

During kick-off meetings and the Intervention Delivery and Evaluation Analysis (IDEA) workshop, all parties worked to develop a detailed theory of change (TOC). The main goals of the meeting were to finalise the EEF's Template for Intervention Description and Replication (TIDieR) framework, and to discuss the PALS Logic Model and how the Implementation and Process Evaluation (IPE) data was to be collected at each stage.[51] The movement to remote learning resulting from the COVID-19 outbreak led to further changes in the proposed implementation and process data collection.

---

[46] Hutchison, D., & Styles, B. (2010). *A guide to running randomised controlled trials for educational researchers*. Slough: NFER.

[47] Higgins, J. P. T., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., Savović, J., Schulz, K. F., Weeks, L., & Sterne, J. A. C. (2011) The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *British Medical Journal, 18*(343), (d5928).

[48] Education Endowment Foundation (2018) *Statistical Guidance for EEF evaluations.* London: Education Endowment Foundation. Retrieved from: https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Protocol_or_SAP/EEF_statistical_analysis_guidance_2018.pdf

[49] Education Endowment Foundation (2018) *Statistical Guidance for EEF evaluations.* London: Education Endowment Foundation. Retrieved from: https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Protocol_or_SAP/EEF_statistical_analysis_guidance_2018.pdf

[50] Allison, P. D., (2012) Handling Missing Data by Maximum Likelihood. Haverford, PA: Statistical Horizons. Retrieved from: https://statisticalhorizons.com/wp-content/uploads/MissingDataByML.pdf

[51] Hoffmann, T. C., Glasziou, P. P., Boutron, I., Milne, R., Perera, R., Moher, D., ... & Lamb, S. E. (2014). Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *Bmj*, *348*, g1687.

The existing literature points to possible barriers to implementation that could be explored in the IPE. Research suggests that PALS may impede pupils' opportunity to learn from errors, that pupils may consider PALS less important than teacher-led activities, and that teachers may relinquish some of their responsibility for teaching.[52] Potential implementation barriers from previous PALS evaluations in kindergarten and first grade classes included lack of fidelity, teacher attrition, and pupil absenteeism.[53,54,55,56] A pilot evaluation in UK found that children were mixing up the activities and were not completing the four activities in the correct order. These findings suggest that it will be important to understand how Year 5 teachers are trained, how RAs and peers support and monitor teachers, how pupils are trained/paired, and how often pairs change, as well as how the partners of absent pupils are treated. Schools can choose to replace guided reading or other scheduled literacy lessons with PALS-UK sessions. They also have the option to deliver PALS-UK supplementary to other literacy lessons. It will be important to study how the programme fits into the school timetable and if there are any supplemental activities that the control schools implement.

The purpose of the process evaluation is to address the following questions:

RQ 1: Was the intervention implemented with fidelity in the intervention classrooms?

RQ 2: What factors and initial conditions appear to explain variation in fidelity of implementation? What are the facilitators and barriers to delivery?

RQ 3: What are the drivers of impact? What are the necessary conditions for success of the programme in terms of achieving impact?

RQ 4: What does "business as usual" look like in control schools?

We have developed a multi-stage mixed-methods IPE data collection plan. Table 1 presents an overview of the activities that will be part of the evaluation alongside the specific aims of these activities. Further details of the data collection methods are discussed in further detail below.

Table 1: Data collection activities and aims

Phase	Activities	Aims
Pre-randomisation	• IDEA workshop	• Development and refinement of TOC (RQ 1, 2, 3, 4)

[52] Richardson, V., & Anders, P. (1998). A view from across the Grand Canyon. Learning Disabilities Quarterly, 21(1), 85-97.
[53] Stein, M. L., Berends, M., Fuchs, D., McMaster, K., Saenz, L., Fuchs, L. S., & Compton, D. L. (2008). Scaling up an early reading program: Relationships among teacher support, fidelity of implementation, and student performance across different sites and years. *Education Evaluation and Policy Analysis, 30*(4), 368-388.
[54] Calhoon, M., Al Otaiba, S., Cihak, D., King, A., & Avalos, A. (2007). Effects of a peer-mediated program on reading skill acquisition for two-way bilingual first-grade classrooms. *Learning Disabilities Quarterly, 30 (*3), 169-184.
[55] Calhoon, M., Al Otaiba, S., Greenberg, D., King, A., & Avalos, A. (2006). Improving reading skills in predominantly Hispanic Title 1 first-grade classrooms: The promise of peer-assisted learning strategies. *Learning Disabilities & Practice, 21*(4), 261-272.
[56] Mathes, P. G., Torgesen, J. K., & Allor, J. H. (2001). The effects of Peer Assisted Literacy Strategies for first-grade readers with and without additional computer-assisted instruction in phonological awareness. *American Educational Research Journal, 38* (2), 371-410.

Phase	Activities	Aims
	• Review of PALS-UK manual and resources	• Documenting the quality of the intervention (RQ 1,2,3)
Delivery period	• Observation of training (initial and top-up training) • Training attendance logs • Surveys of head-teachers in treatment and control arm • Review of the observation fidelity checklist	• Documenting the quality and teacher responsiveness to intervention of the intervention (RQ 1, 2) • Documenting intervention compliance (RQ 1,2) • Documenting motivation for joining the programme; Documenting counterfactual and any interventions that may impact on child reading attainment (RQ 1,2,4) • Documenting dosage and the quality and intervention compliance (RQ 1, 2, 3)
Post-delivery period	• Headteacher survey in treatment arm • Review of pupil pairing logs • Headteacher survey in control arm • Teacher survey in treatment arm • Phone interviews with teachers, peer observers, head teachers	• Exploring experience of settings (RQ 2); Reporting on costs associated with delivering PALS (cost section of final report) • Documenting quality of pupil pairing (RQ 1,2,3) • Documenting counterfactual and any interventions that may impact on child behaviour or language (RQ, 4) • Exploring experience of teachers and teacher responsiveness to intervention (RQ 1, 2, 3,) • Document the quality of intervention and exploring perceptions of programme impact and drivers of change (RQ 1, 2, 3)
Contingency planning activities resulting from COVID-19	Headteacher survey in treatment arm Headteacher survey in control arm Teacher survey in treatment arm	• Exploring experience of settings (RQ 2) • Documenting counterfactual and any interventions that may impact on child behaviour or language (RQ, 4)

Phase	Activities	Aims
	Phone interviews with teachers, head teachers	• Exploring experience of teachers (RQ 1, 2, 3,) • Document the quality of intervention and exploring programme impact and drivers of change (RQ 1, 2, 3)

Pre-intervention phase

Review of the manual

The intervention manual wasfinalised and shared with RAND Europe before the training begun. RAND has reviewed the manual to build better understanding of the programme and its key components, with the aim to inform the later IPE tasks, such as the observations and the phrasing of wording of survey/interview questions.

During Intervention phase

Observation of training of Year 5 teachers

From each school, one **Year 5 teacher and one peer observer** attended two days of training. Training sessions were arranged as efficiently as possible, with around twenty schools and forty teachers attending each training session. RAND Europe observed the initial and top-up training of Year 5 teachers to gather information about how the training is delivered, including teacher engagement with the training they receive, as this may affect how they have delivered sessions and, consequently, the students' outcomes. This was recorded in the form of contemporaneous, structured note-taking rather than video recording.

Attendance at training

Attendance at training is a key part of the TOC in terms of leading to better outcomes. If teachers in the intervention group have not attended training, it is unlikely they will be able to deliver the intervention. As such, training is considered mandatory. Teachers who missed training were encouraged to attend an alternate date the same week. If they are unable to attend an alternate date, the delivery team provided appropriate training on a case by case basis, by visits to individual teachers within the following week. It is important to know whether teachers responsible for delivering the programme have or have not attended. By submitting MoUs, schools have committed to sending a teacher and peer observer to the training, so this is not anticipated to be a substantial problem. However, in the survey administered to Year 5 teachers, we asked Year 5 teachers themselves to self-report their attendance at training, as well as to provide their perceptions on this training.

Surveys

The surveys collect data on usual practices, attitudes, perceptions and reading related activities in the classroom from headteachers, teachers, and peer observers. Surveys are

developed by the project team at RAND Europe in consultation with the delivery team and the EEF. The questions are tailored to each type of respondent in each group (intervention and control). Surveys are kept as short as possible and we expect that it takes no more than 15 minutes to complete each. All surveys will be uploaded onto the SmartSurvey platform by RAND Europe and completed online. Survey data has been collected at three different time points. Headteachers from intervention and control schools will be surveyed at baseline in October, and post-intervention in May and in October. In addition to headteachers, teachers will be surveyed in intervention schools in May 2020, and in October 2020.

Headteacher Survey 1

The main objective of the first baseline survey with headteachers was to examine the motivations for joining the trial and their understanding of the intervention, so that potential barriers to recruitment can be better understood. The survey also asked about any other reading interventions that may be in use in the school and about business as usual more generally. Survey data from Head Teachers at baseline was collected in Autumn 2019. This was scheduled after randomisation and before programme delivery. Therefore, it offers an opportunity to ask headteachers how they feel about their allocation.

Case study schools selection

Four intervention schools are studied in greater depth as purposive case studies to give further depth of information on the implementation of PALS-UK in the intervention schools. Schools were selected randomly for case studies on the basis of a list of schools with varying leadership's engagement and enthusiasm for the project, as preliminarily reported by the delivery team. We also considered region and school size when selecting in order to allow schools from different regions and with a range of levels of engagement with the programme to be studied in more depth. We have been conducting semi-structured interviews with headteachers, teachers and peer observers in case study schools with questions designed to find out more about perceptions and experiences regarding implementation of PALS-UK. We will draw upon existing observation data from the RA and peer classroom observations in order to evaluate the quality of pupil pairing and assess children's perspectives of PALS-UK.

Review of RA and peer classroom observation

Classroom observations were completed in all schools allocated to the intervention. Two observations were completed by an RA (once during the initial four week training and once in the first four weeks of the intervention) and two more by a peer observer teacher from the school (twice after the initial RA observations). The delivery team run training on how to do the observation session. During the IDEA workshop it was agreed to simplify the observation sheet for the peer observer by asking them to observe one pair only. The observers used a checklist to assess how well PALS was implemented in the classroom. RAs completed one observation in the first four weeks and another in the following four weeks. Peer teacher observations were scheduled for week seven and twelve of programme delivery. The observation checklist score will be used in the construct of the compliance measure. Additionally, we will explore differences in the checklist score for case study schools.

Post-intervention phase

Headteacher Survey 2

The second headteacher survey will help us determine if control schools have changed their literacy practices in the last school year. We have asked headteachers in intervention schools if they believe practices in teaching reading have changed as a result of PALS-UK, and if so, how. This will allow us to corroborate information obtained from teacher surveys. The survey will also allow us to understand the potential costs associated with the intervention through asking about additional support time/costs.

Survey of Year 5 teachers in intervention schools

In May 2020, following the programme implementation, we have surveyed Year 5 teachers and peer observers in intervention schools only about their experience of delivering the intervention. The survey included questions on the following:

- **Perception of training**. In the logic model, the training provided is the primary means for those delivering the intervention to develop both the knowledge and pedagogical skills needed. This part of the survey asked teachers how useful they believe the training has been in preparing them to deliver the intervention.

- **Perception of programme effect**. The survey included questions about the effect of PALS-UK on peer relationships and whether the strength of these relationships was affected by the intervention. It will also ask about the usefulness of each element of the programme: partner reading, re-tell, paragraph shrinking, prediction relay.

- **Use of online resources/ongoing support.** The survey will probe teachers about their use of the resources and support in order to gauge the level of take-up as well as how useful these resources and support have been in helping teachers deliver the intervention effectively.

- **Perception of the usefulness of the PALS-UK manuals and other resources (e.g. videos and gap tasks).** This information is important to determine if amendments to the resources provided are needed in the future.

- **Support from RAs to Year 5 teachers**. The survey will ask Year 5 teachers if support from RAs has been available.

- **Post-trial implementation**. The survey will ask whether teachers continued or will continue to implement PALS after the twenty weeks of the efficacy trial.

Interviews with Headteachers, Teachers and Peer observers in case study schools

Teachers, peer observers and headteachers in the case study schools have been invited for semi-structured interviews in order for the evaluation team to obtain a more detailed understanding of their experiences in the trial. In particular, interviews with those who have delivered the intervention will also help determine the significance of the drivers of change that have been identified in the logic model and what 'on the ground' practice looks like. Questions focus on the overall experience of implementing PALS-UK and any barriers to it. The interviews will help us find out whether teachers have supported/sustained the intervention. It

will allow us to gain detailed information of teachers' perceptions of how the programme impacted the way teachers work and how they felt it impacted pupils.

After a teacher, peer observer or headteacher has agreed to participate in an interview, then a RAND Europe researcher has arranged a date and time to conduct the interview over the phone. The interviews lasts between 30 and 40 minutes and are recorded but not transcribed. Telephone interviews are preferred here because they are more flexible and cost-effective. Interviewers have been taking contemporaneous notes but only anonymised quotes will be reported. Participants will be interviewed twice: towards the end of the intervention in Summer 2020 and following the start of the new school year in Autumn 2020.

Children's perspectives

It will also be important to find to what extent children understand the different tasks in PALS-UK. This will give us an insight in how well the teachers have trained the pupils in the PALS-UK approach. RAs from the delivery team (Coventry University or NTU) asked children during the observation what each activity is called and what to do for each activity. RAND Europe will analyse this data only for the case-study schools, where such data can be properly contextualised and analysed within the context of greater knowledge of the school. If the data is available systematically then we will explore whether this information can be incorporated into the statistical analysis.

Pupil pairing

The evaluation team will review the pairing logs filled by teachers for the schools selected as case studies. The quality of the pairs will be assessed by comparing the pairing logs with the rankings based on the PIRA assessment. Relatedly, RAND will review how often pairs change, as well as how the partners of absent pupils are treated.

UPDATE: COVID 19 driven phase

We would like to understand the activities of schools in both the intervention and the control conditions since the end of PALS-UK delivery, and their experiences with literacy/reading support and interventions since the start of the lockdown and social distancing rules. Therefore, we will run another round of IPE activity that is similar to the original endline IPE activity, to include:

1) A survey of head teachers in intervention schools, asking about PALS-UK use since the start of the lockdown and social distancing, with a focus on reading interventions/support/catch-up;

2) A survey of head teachers in control schools, asking about general reading and literacy work since the start of the lockdown and social distancing;

3) A survey of teachers in intervention schools, asking about PALS-UK use since the start of the lockdown and social distancing, with a focus on reading interventions/support/catch-up, and also their experiences working with children since the return to schools;

4) A set of four case studies in intervention schools, using interviews with head teachers and teachers, to understand in more depth any interventions/support/catch-up related to reading in the period in-between the end of the intervention and the outcome testing, to further contextualise the findings from the trial.

IPE Data analysis

The rich data from the IPE will be summarised to provide a description of how the intervention worked in practice. Descriptive findings from the survey and interviews will be aggregated and summarised, with results used to provide a description of how the intervention worked in practice. Additionally, any open-text responses will be analysed using a general inductive approach, with results used to provide a description of how the intervention worked in practice.[57] In addition, this data will be used to empirically examine the assumptions underlying the key mechanisms and processes in the PALS logic model, exploring how they worked in the context of this evaluation. This modelling is considered a key way of understanding the association between activities and outputs to their intended outcomes and is considered particularly important when the change in distal outcomes (in this case, reading attainment) is assumed to be underpinned in some way by direct effects on proximal outcomes (in this case, reading fluency and self-efficacy).[58]

A description of the counterfactual will also be provided based on the responses from practitioner surveys in control settings.

Questions in the surveys and interviews will be designed with reference to the EEF's IPE handbook and will collect data on both financial costs and practitioner time.

Compliance measure

To enable a non-compliance analysis, compliance will be defined at the school level, based on completion of programme activities, as recorded by the Delivery team. The compliance measure is discussed in the Statistical Analysis Plan.

Cost evaluation

As discussed above, the main costs of the intervention relate to training, materials, and teachers' time spent training and supporting the delivery of the programme. In order to calculate the cost of training and materials, the evaluation team (RAND Europe) will rely on data provided by the Delivery Team (Nottingham Trent University and Coventry University). RAND Europe will also take into account the cost of teachers' time spent delivering the programme. We will also gather cost data through the surveys and interviews in the implementation and process evaluation (see above on pp. 25-26).

Questions in the surveys and interviews were focused on assessing any pre-requisite costs (such as training costs and materials) and any direct and marginal costs that are directly attributable to schools' participation in the intervention (for example, staff time). We will use this information to estimate cost per-pupil, following EEF guidelines on cost evaluations.[59]

[57] Thomas, D. R. (2006). A General Inductive Approach for Analyzing Qualitative Evaluation Data. American Journal of Evaluation, 27(2): 237 – 246.
[58] Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R., & Kerr, K. (2016). Implementation and process evaluation (IPE) for interventions in education settings: An introductory handbook. Education Endowment Foundation, 1.
[59] Education Endowment Foundation (2015)*EEF Guidance on Cost Evaluations.* London: Education Endowment Foundation. Retrieved from: https://v1.educationendowmentfoundation.org.uk/uploads/pdf/EEF_guidance_to_evaluators_on_cost_evaluation.pdf

Ethics and registration

The trial has been registered on the ISRCTN registry, which stands for 'International Standard Randomised Controlled Trial Number' and is used to describe RCTs and efficacy trials at inception. The trial has been assigned an ID registration number: ISRCTN10664882.[60]

The ethics and registration processes are in accordance with the ethics policies adopted by RAND Europe Ethical approval for the intervention was granted by the Coventry University Ethics Committee. The reference number for this approval is ED-CIA-18-192, while the ethics projects numbers are P90795 (parents and pupils) and P86212 (Schools and class teachers). The RAND U.S. Human Subjects Protection Committee (HSPC) has approved the study with reference number IRB00000051.

Parents may withdraw children from the trial. Parents or legal guardians act as decision-makers for individual pupils. Prior to pupil data being sent to the Delivery team, parents will be sent information and withdrawal forms by the school and have the opportunity to return these. The parental information sheets and withdrawal forms will be sent out to parents by the schools after the school representatives sign the Memorandum of Understanding describing what is involved in the trial. Parents can withdraw their children at any time from the research, but the initial withdrawal forms will be requested to be returned by parents within 2 weeks. If participants choose to withdraw their children from the study later on, their data will not be collected and will be deleted as appropriate (see privacy notice).

RAND Europe will collect consent forms for headteachers, teachers, and other school staff who will volunteer to participate in an interview. The cover page for each survey will contain an informed consent and data protection statement for respondents. This will inform respondents that participation in the survey is entirely voluntary.

Any data sharing required will be governed by the data sharing agreement signed between the funder (the EEF), the Delivery team (Coventry University) and the Evaluation team (RAND Europe).

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

Data protection

RAND Europe will obtain personal data from schools and pupils as a data controller. Basic pupil information will be obtained on the basis of legitimate interests from schools pursuant to brief data sharing undertakings or agreements with each school that is recruited. RAND Europe shall obtain baseline data under a data sharing agreement with Coventry University. RAND Europe shall obtain pupil outcome data using its own testing force, which will be trained by Coventry University. Outcome data will be on the basis of legitimate interests. Pupils and parents shall be provided with age-appropriate fair processing privacy notices that explain the use, storage and secure handling of the data. This will also include an option to withdraw of the study.

[60] The study can be viewed at: http://www.isrctn.com/ISRCTN10664882

Data will only be saved on GDPR compliant, secure servers inside the EEA or UK. All processes will be handled in accordance with RAND's Data Protection Policy. RAND Europe is registered with Information Commissioner's Office (ICO), registration number Z6947026 and is certified for adhering to ISO 9001:2015 quality management practices. In order to stratify the sample and adequately evaluate the intervention as outlined in this proposal, it is necessary to process special categories of data, namely FSM status of pupils. RAND Europe considers this endeavour to fall under GDPR, Chapter 2, Article 9, Paragraphs 2d) and 2g).

RAND will delete all data one year after the project ends. For the purpose of research, following the completion of the trial, the data will be shared with the EEF archive, at which point EEF will become the data controller.[61] The data will be shared with the Department for Education, the EEF's archive manager and, in an anonymised form, with the Office for National Statistics and potentially other research teams. Further matching to NPD and other administrative data including KS2 may take place during subsequent research.

Personnel

Delivery team: PALS-UK (Coventry University and Nottingham Trent University)

Project Leader(s): Dr. Emma Vardy (Nottingham Trent University) and Dr Helen Breadmore (Coventry University)

Evaluation team: RAND Europe

Project Leader: Dr Sashka Dimova (RAND Europe)

Project Manager: Dr Andreas Culora (RAND Europe)

Core fieldwork and analysis team: Dr Andreas Culora, Lucy Gilder, Annemari de Silva (RAND Europe)

Risks

Risk	Assessment	Mitigation strategy
Recruitment failure	Likelihood: Low Impact: High	Recruitment progressing well.
Attrition	Likelihood: Moderate Impact: Moderate to high	Clear information about expectations and requirements provided to participating schools. MoU to be signed with participating schools Intention to Treat (ITT) analysis to be used. Attrition to be monitored and reported according to CONSORT guidelines.[62]

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

[62] Campbell, M. K., Piaggio, G., Elbourne, D. R., & Altman, D. G. (2012). Consort 2010 statement: extension to cluster randomised trials. *Bmj*, *345*, e5661.

		Schools in control group will receive their payment for participating in the trial after outcome testing has been completed. This is an incentive to remain in the trial.
Different rates of attrition from control and treatment groups	Likelihood: Low Impact: Moderate	There is a risk that schools in the treatment group may face an extra burden in terms of time and resources to deliver the programme. This can be mitigated by regular liaison with schools to secure continued engagement in the trial. Schools would have agreed to the terms of the MoUs, which include the commitment for data to be collected at various stages.
Missing data	Likelihood: Moderate Impact: Moderate	To limit the amount of missing data, baseline and post– trial outcome testing will happen in an extended period (approximately a month).
Pupil mobility	Likelihood: Moderate Impact: Low	Pupils who are included in the study at the start of the school year and who move between study schools will be retained and analysed according to their original allocation to treatment/ control. Pupils who migrate to non-study schools will be excluded from the analysis as testing is via external tests (rather than via Key Stage 2 data). In the event that mobility to non-study schools exceeds 10 per cent on average across all schools, then the evaluators will discuss with the EEF the possibility of additional funding to collect this information.
Low implementation fidelity	Likelihood: Low to moderate Impact: Moderate	Process evaluation to monitor and document fidelity of implementation.
Cross-contamination	Likelihood: Low Impact: Moderate	Clear instructions will be provided to participants about the trial to avoid contamination. It can happen if single pupils move schools allocated to different groups.
Evaluation team members absence or turn-over	Likelihood: Moderate Impact: Low	All RAND staff have a three month notice period to allow sufficient time for handover. The team can be supplemented by researchers with experience in evaluation from the larger RAND Europe pool.
Low response rates for online surveys	Likelihood: Moderate Impact: Moderate	Online surveys to be kept to a maximum of 15 minutes. Respondents given the opportunity to complete survey online on multiple occasions if required. Sufficient data collection window given with real-time monitoring of response rates to allow for reminders to be targeted.
Lack of coordination with larger teams	Likelihood: Moderate Impact: Moderate	Teams to attend initial meetings and agree on roles and responsibilities at the outset. Regular updates to be provided to the lead evaluators. Regular contact between senior team from each organisation.

		Single point of contact for RAND team in terms of project management.
Further disruptions to trial activities due to COVID-19 (or alike) pandemic	Likelihood: Moderate Impact: High	A new outbreak of the COVID-19 pandemic could further delay evaluation activities. Teams will maintain regular contact to coordinate and decide on any strategies to mitigate potential risks to the evaluation plans.

Timeline

Dates	Activity	Staff responsible/ leading
January 2019	IDEA workshop	RAND Europe
February- June 2019	Recruiting schools and teachers	Nottingham Trent University/ Coventry University
June- July 2019	Withdrawal forms to be sent to parents	Schools
July 2019	Collection of pupil information for all Year 4 pupils	Schools
July- August 2019	School and pupil information to be collected sent to RAND	Nottingham Trent University/ Coventry University
August- September 2019	Selection of Year 5 class to undertake outcome testing in multiple entry form schools	RAND Europe
September 2019	Baseline outcome testing	Nottingham Trent University/ Coventry University
September 2019	Randomisation	RAND Europe
September 2019	Notifying schools of allocation	Nottingham Trent University/ Coventry University
September 2019	Initial training for Year 5 teachers	Nottingham Trent University/ Coventry University
October 2019	Baseline survey of Headteachers all schools	RAND Europe
November 2019	Completion of Statistical Analysis Plan	RAND Europe
November 2019	Top-up training	Nottingham Trent University/ Coventry University

October 2019-April 2020	Programme implementation	Nottingham Trent University/ Coventry University
September 2019-April 2020	Compilation of training attendance records, combined with observation checklist score	Nottingham Trent University/ Coventry University
May- November 2020	Collection of surveys from Year 5 teacher and Headteachers	RAND Europe
November - December 2020	Outcome testing	RAND Europe
November 2020-April 2021	Interviews in case study schools (Year 5 teachers, peer observers, Headteachers)	RAND Europe
May 2021	Submission of draft EEF Report	RAND Europe
December 2021	Final EEF Report	RAND Europe

Update: Note that we have updated the timeline to reflect changes in activities and disruptions caused by Covid-19