Peer Assisted Learning Strategies - UK

Evaluation Report

February 2022

Andreas Culora, Sashka Dimova, Sonia Ilie, Alex Sutherland, Lucy Gilder

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:

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 2

Executive Summary 3

Introduction 5

Methods 13

Impact evaluation 34

Implementation and process evaluation 59

Cost 93

Conclusion 97

References 103

Appendix A: EEF cost rating 105

Appendix B: Security classification of trial findings 106

Appendix C: Recruitment Documentation 108

Appendix D: Baseline equivalence 127

Appendix E: Distribution of PiRA scores in the final sample by treatment 131

Appendix F: PALS-UK reading self-efficacy questionnaire 138

Appendix G: Randomisation code, analysis code and output 153

Appendix H: Implementation and Process Evaluation Privacy Notices

Appendix I: Secondary outcome - Reading comprehension (WIAT-III-UK-T)

Appendix J: Full observation data

About the evaluator

The project was independently evaluated by a team from RAND Europe: Andreas Culora, Sashka Dimova, Sonia Ilie, Alex Sutherland, and Lucy Gilder.

The lead evaluator was Sashka Dimova.

Contact details:

Name: Dr Emma Disley

Address: RAND Europe, Westbrook Centre, Milton Road, Cambridge, CB4 1YG, UK

Tel.: +44 1223 353 329

Email: edisley@randeurope.org

Executive summary

The project

Peer Assisted Learning Strategies UK (PALS-UK) is a whole-class, structured peer tutoring intervention aiming to improve pupils' reading comprehension, oral reading fluency, and overall reading attainment. Pupils are paired and take part in four activities: partner reading (pupils take turns reading while their partner corrects any reading errors), re-tell (one pupil retells the passage just read with prompts from their partner), paragraph shrinking (each pupils reads a short passage and summarises what they have read), and prediction relay (pupils take turns to read and predict what will happen in the next half-page of the story). PALS was developed in the United States by Douglas Fuchs, Lynn Fuchs, and other researchers from Vanderbilt University. The intervention was adapted for the U.K. context and delivered by Dr Emma Vardy (Nottingham Trent University) and Dr Helen Breadmore (University of Birmingham). The U.K. delivery team members were affiliated to Coventry University at the outset of the trial. The trial started in February 2019 and ended in April 2021, but was disrupted by the COVID-19 pandemic.

The programme was delivered by teachers who received one day of initial training from the PALS-UK team followed by a half-day top-up training. Pupils also received training from their teacher over the first four weeks of the programme to learn how to undertake the four PALS activities and work collaboratively with their peers. This project was delivered to Year 5 pupils (age 9 to 10) three times a week for 30-35 minutes per session for 20 weeks.

The efficacy trial was a two-arm stratified cluster-randomised controlled trial, randomised at the school-level; 89 schools and 2,283 pupils took part. The trial tested the impact of PALS-UK on pupils' reading attainment, measured by the PIRA assessment (the primary outcome), as well as on oral reading fluency, reading comprehension, and reading self-efficacy (secondary outcomes). An Implementation and Process Evaluation (IPE) was also conducted, including observations of training and delivery, surveys, and interviews with school staff. Project delivery took place as planned but due to the COVID-19 pandemic, primary outcome testing was delayed by six months; secondary outcome testing for reading fluency and comprehension could not be completed and IPE activities were disrupted.

Table 1: Key conclusions

Key conclusions	
1.	Children in PALS-UK schools made the equivalent of one month less progress in reading, on average, compared to children in other schools. This result has a low EEF padlock security rating. Given the impacts of the COVID-19 pandemic on the trial, there is uncertainty about whether this trial can provide a generalisable estimate of the impact of PALS-UK.
2.	Children in PALS-UK schools who were eligible for free school meals made the equivalent of one months' additional progress in reading, on average, compared to children in the control group eligible for free school meals. This result has less security than the overall findings and is similarly affected by the impacts of COVID-19 on the trial and the low overall security rating.
3.	There was no evidence that PALS-UK impacted pupils' reading self-efficacy, based on pupil responses to a self-efficacy questionnaire.
4.	Teachers surveyed and interviewed as part of the process evaluation felt that PALS-UK had positive effects on pupils' reading confidence and reading fluency and, to a lesser extent, on reading attainment.
5.	Observations of programme delivery suggest that most schools implemented PALS-UK with excellent fidelity. Survey data also indicates that 100% of schools delivered PALS-UK for the full 20-week period and over 95% of teachers attended the initial and top-up training, although low survey response rates could have biased the results. These findings suggest that the programme was delivered well from October 2019 to March 2020, although due to COVID-19 disruption, schools could not continue to deliver PALS-UK as intended between March 2020 and the delayed start to outcome testing in November and December 2020.

EEF security rating

These findings have a low security rating. This was an efficacy trial, which tested whether the intervention worked under developer-led conditions in 89 schools. The trial was a well-designed two-armed randomised controlled trial and was well-powered. However, 36% of the pupils who started the trial were not included the final analysis due to schools withdrawing from the trial and pupil absences. Most school withdrawals and pupil absences are likely to have been due to challenges associated with the COVID-19 pandemic, including staff and pupil illness or self-isolation, school workforce capacity issues, and the shift to remote learning. Significant differences between the intervention and control group in numbers of pupils not included in the final analysis make it harder to accurately estimate the size of the impact on the pupils in the trial.

Additional findings

The results of this trial should not be interpreted as conclusive evidence of the effects of PALS-UK on pupils' reading. The COVID-19 disruption limits our confidence in the generalisability of the findings. Pupils in PALS-UK schools made, on average, one month less progress than those in the control group equivalent. This is our best estimate of impact, which has a low security rating. As with any study, there is some additional statistical uncertainty around the result: the possible impact of this programme ranges from negative effects of four months less progress to positive effects of up to two months additional progress.

This evaluation was designed to capture the immediate effect of PALS-UK on Year 5 pupils' reading attainment. Due to the COVID-19 pandemic, outcome testing that was initially scheduled for May and June 2020 was delayed to November and December 2020. Schools could not continue to deliver PALS-UK as intended (if at all) after March 2020 and school staff reported that they perceived the COVID-19 pandemic to have had a negative impact on pupils' reading skills during this period. It is possible that the delay in post-testing, combined with these contextual factors, may have diluted any impacts of PALS-UK over time making it harder to detect the true effect of the intervention on pupils' reading.

Surveyed and interviewed teachers reported that the training, programme resources, and ongoing support were high quality and useful. There was less consensus around the usefulness of the pupil training for informing pupils how to work well in pairs and the usefulness of the top-up training for informing the selection of appropriate reading books. Staff interviewed were more positive about the perceived impact of PALS-UK on reading confidence and reading fluency than on the impact on reading comprehension. Staff also felt that PALS-UK had positive impacts on teachers' reading instruction, pupils' vocabulary, and pupils' independent learning.

Wider research suggests that peer assisted learning approaches can have positive impacts on a range of outcomes. This evaluation was aiming to provide robust evidence on the effectiveness of PALS as a peer assisted learning approach in the U.K. context. However, due to disruption to this trial associated with the Covid-19 pandemic and high levels of statistical uncertainty surrounding the results, this study is unable to provide robust evidence of the effect of PALS-UK on pupil outcomes.

Cost

The average cost of PALS-UK for one class was around £181, or £2.57 per pupil per year when averaged over three years. A minority of schools purchased additional materials, such as books, to support delivery and consequently incurred slightly higher costs. Additionally, most schools needed to provide two days of staff cover for teachers to attend the training days.

Impact

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
Reading attainment (age standardised PiRA score)	-0.07 (-0.29,0.14)	-1		1,459	0.478	£ £ £ £ £
Reading attainment, FSM pupils (age standardised PiRA score)	0.05 (-0.16, 0.27)	+1	n/a	1,451	0.630	£ £ £ £ £

Introduction

Background

Peer tutoring interventions are appealing due to their potential impact and low cost. Academic research and the EEF Teaching and Evidence Toolkit (EEF, 2018a) reports evidence of moderate to high effects for peer tutoring, as well as cross-age tutoring (Topping et al., 2011; Lloyd et al., 2015; Lloyd et al., 2015b; Shenderovich et al., 2016). Although there is a long history of peer tutoring interventions in the U.K., results from recent empirical studies have not been encouraging (Lloyd et al., 2015).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. One explanation for the different results is that peer tutoring appears to be more effective when highly structured so implementation and fidelity will matter, as might subgroups. For example, an RCT in 80 schools evaluating paired reading in Scotland identified larger effects for girls, pupils from low SES groups, and those with lower reading ability (Topping et al., 2011). In the U.K., the evidence is stronger for cross-age than same-age tutoring (Tymms et al., 2011). Accordingly, Peer Assisted Learning Strategies U.K. (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 is a structured peer reading programme originally developed in the United States by Douglas Fuchs, Lynn Fuchs, and other researchers from Vanderbilt University, with the main aim to improve oral reading fluency and reading comprehension in the Grade 2–6 version. PALS has been widely researched in the U.S. but a What Works Clearinghouse (WWC) effectiveness summary for 'beginning readers' found that only three studies met WWC evidence standards. 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 two studies, focusing on nine- and ten-year-old pupils who received 15 weeks of PALS that met their evidence standards (WWC, 2012). The summary concluded that PALS demonstrated potentially positive effects on alphabetics, no conclusive effects on reading fluency and mixed effects on reading comprehension. The findings around reading comprehension were mixed as one study found 'substantively important negative effects' of PALS, while the other found 'substantively important positive effects' of the intervention in this domain (ibid.).

Given the existing evidence, there is substantial value in evaluating PALS in the U.K. Additionally, as PALS-UK is one of the few programmes that focuses on fluency as a key component in reading and in learning to read, this study therefore aims to provide valuable evidence on the nature of effective instruction in fluency. PALS-UK would also benefit from an independent evaluation as previous evaluations of PALS have been developer-led. The overall aim of this EEF-funded efficacy trial is to help determine whether PALS-UK leads to observable improved outcomes for pupils.

Evaluation overview

The evaluation was designed and executed as a stratified, two-arm, cluster-randomised controlled trial (cRCT), across 89 schools from two geographical regions (the Midlands and the North East). Of the 89 schools recruited to PALS-UK, the programme was delivered in 44 intervention schools with the remaining 45 schools assigned to a business as usual (BAU) control condition. In this evaluation, PALS-UK was delivered to Year 5 children.[1] This evaluation of PALS-UK was comprised of three strands: (1) an impact evaluation, to assess whether PALS-UK led to an observable improvement in outcomes for Year 5 pupils, (2) an implementation and process evaluation (IPE) which provided contextual information around programme delivery and the 'business as usual' scenario during the trial, and (3) a cost evaluation, which provided information on the delivery costs incurred by intervention schools due to delivering PALS-UK. The movement to remote learning as a result of the COVID-19 pandemic in March to July 2020 and in January 2021 effected the outcome data collection activities and consequently the extent to which we can be confident in our findings. As a consequence, analysis on one of the secondary outcomes—the Wechsler Individual Achievement Test–III U.K. for Teachers (WIAT-III-UK-T)—was not feasible. Further details around the trial design changes can be found in the updated evaluation protocol (Dimova et al., 2021).

[1] In some schools, PALS-UK was also delivered to pupils in other year groups. However, data was only collected from Year 5 pupils in this evaluation.

The evaluation was funded by the Education Endowment Foundation (EEF). The programme was delivered by the PALS-UK delivery team (Nottingham Trent University and Coventry University) and the trial was carried out by RAND Europe; the independent evaluator appointed by the EEF.

Intervention

Figure 1 sets out the implementation logic model developed during the trial set-up meetings and the IDEA workshop between the EEF, the PALS-UK delivery team, and RAND Europe (the evaluation team). The logic model and TIDieR checklist provide structure for the rest of this section.[2]

The implementation logic model (Figure 1) for PALS-UK involves a mandatory element of initial and top-up training. In addition to training, Year 5 teachers were offered support during implementation on an ad hoc basis. Training and appropriate support are key prerequisites for successful delivery. These enable teachers to successfully train pupils on PALS, and to successfully deliver PALS sessions in their classroom. As can be seen in the stated logic model, PALS-UK aims to improve pupils' confidence in reading, oral reading fluency, and reading comprehension. The logic model considers improvements in those reading related skills as essential elements for improved reading attainment.

We will now use the TIDieR checklist to describe the PALS-UK intervention in more detail.

Why

Peer Assisted Learning Strategies for Reading UK (PALS-UK) is a whole-class, structured, paired reading intervention that was designed to improve children's oral reading fluency and reading comprehension skills. The sessions focused on improving pupils' oral reading fluency and comprehension as teaching children to read fluently and comprehend texts is fundamental for school achievement.

Each PALS-UK session consists of four different activities—partner reading, re-tell, paragraph shrinking, and prediction relay—that are designed to improve oral reading fluency or reading comprehension. Partner reading is designed to improve oral reading fluency. Re-tell, paragraph shrinking, and prediction relay are designed to improve reading comprehension. Pupils used the PALS 'check-it' procedure throughout these activities to check for errors made by their partner and provide feedback. In addition to oral reading fluency and reading comprehension, it is expected that increasing the practice of reading aloud and completing reading comprehension activities with structured peer support will help pupils' reading self-efficacy. Being both the coach and reader was expected to develop confidence levels for both the weaker and stronger reader. In theory, PALS-UK was expected to improve oral reading fluency and comprehension, which would influence both reading self-efficacy and peer relationships. In the long term, oral reading fluency and reading comprehension were expected to influence overall reading attainment.

Who (recipients)

The programme was targeted at Year 5 pupils. In some schools, PALS-UK was also delivered to pupils in other year groups, however, data was only collected from Year 5 pupils in this evaluation.

What (physical or informational materials were used in the intervention)

The intervention was delivered using resources developed by Nottingham Trent University and Coventry University with the support of programme developers from the United States. All Year 5 teachers in the intervention schools received the PALS-UK manual and a selection of 20 reading books for children that are needed to deliver the sessions.

What (procedures, activities are being used)

Prior to PALS-UK implementation, all Year 5 teachers participated in training. The PALS-UK training comprised:

1. An initial one-day training (mandatory). The full day initial training took place prior to implementation. It provided teachers with background information about the skills that underpin reading comprehension and a detailed

[2] For more information, please see here https://www.equator-network.org/wp-content/uploads/2014/03/TIDieR-Checklist-PDF.pdf

description of the intervention. It enabled teachers to practise the PALS activities and to understand the process of pairing pupils. The delivery team shared tips and hints to support the implementation of PALS sessions.

2. A top-up training (mandatory). The top-up training took place after teachers had delivered the four weeks of whole class training. The top-up training focused on book selection, how to change pairs, and sharing good practice.

3. Ongoing support to implementation (optional). During implementation, the delivery team maintained an ongoing relationship with participating schools by providing support in the form of:

    a. videos of how to deliver PALS-UK activities were shared with teachers;

    b. emails to teachers with reminders to switch pairs and tips for book selection; and

    c. additional support offered by research assistants (RAs) to teachers, where teachers could get in touch with an RA via email for additional, specific support.

4. Pupil training (mandatory). The training for children lasted for four weeks and was provided by their Year 5 teachers. Children learnt how to deliver the activities and how to work well in pairs. This training formed part of the PALS-UK sessions, with a new skill introduced during each session. Each lesson was fully manualised and all resources are provided in the PALS-UK manual.

PALS-UK sessions encouraged structured work in *pupil pairs*. Pupils were arranged by their teachers into pairs using available pupil reading-ability data. Teachers in the trial were advised to rank pupils based on ability in reading and then to split the group by the median of the class. Then, the strongest reader of the above-median group was paired with the strongest reader of the below-median group. Thus, there was some gap in reading ability between the pupils but ideally one that was not too large. Teachers were able to reorganise pupils if they felt that the two pupils might not work well together. Pupil pairings were displayed on a board so the pupils knew who it was that they were paired with. Additionally, teachers recorded the pairings in the classroom and selected the reading material for each pair, based on the children's reading ability. Pupils with severe Special Educational Needs and Disability (SEND) who required the support of a full-time teaching assistant (TA) could be paired with their TA.

Pupils swapped pair every four weeks. Furthermore, teachers were advised to re-pair students if something did not work well. Instructions on how to pair children were included in the manual. Additionally, during the top-up training the pupil pairing and re-paring process was discussed in detail as this is core element of the programme.

PALS-UK sessions comprised of four activities: (1) *partner reading*, in which the coach corrects the reader's reading errors using the PALS-UK 'check-it' procedure, (2) *re-tell*, in which the reader re-tells a story with prompts from the coach, (3) *paragraph shrinking*, in which both pupils summarise what they have read using up to ten words, and (4) *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. Sessions followed the same structure with these four activities. Books and appropriate reading materials for the activities were chosen by the teachers and they were selected to fit the reading ability of the pupil with lower prior reading attainment.

How

PALS-UK sessions were delivered in the classroom.

Training for school staff was delivered face-to-face. The initial and top-up training were delivered in a range of locations and teachers could decide which they attended.

When and how much

PALS-UK was delivered over 20 weeks in the 2019/2020 academic year. Each week, teachers delivered three sessions lasting between 30 and 35 minutes each.

It was recommended that PALS-UK sessions should replace guided reading or other supplementary literacy activities (such as whole class reading) that a school normally delivers, but this was subject to school preference.

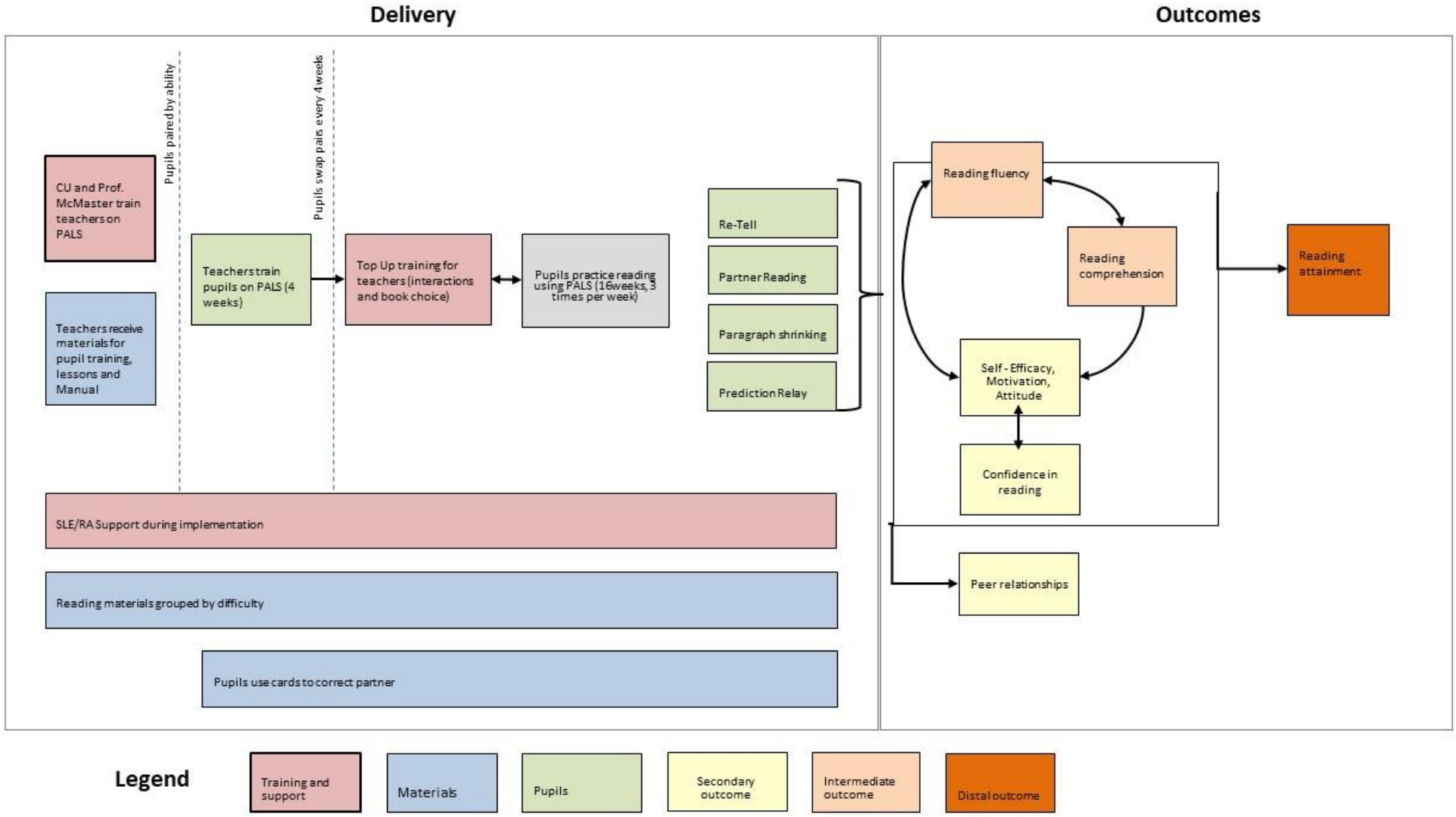
Who (providers, implementers)

PALS was a programme originally developed in the United States by Douglas Fuchs, Lynn Fuchs, and other researchers from Vanderbilt University. The full-day initial training was delivered by Prof. Kristen McMaster (a PALS trainer from the United States) and the PALS-UK delivery team (Dr Emma Vardy and Dr Helen Breadmore). The PALS-UK delivery team provided the top-up training and ongoing support to schools while Year 5 teachers were delivering PALS-UK sessions to their pupils.

Evidence of implementation variability

At the school level the intervention was delivered as planned with no significant variation across treatment schools. The only exception related to schools for which we do not have information as they did not want to take part in data collection activities, which were delayed due to the COVID-19 pandemic. Compliance was assessed as part of this trial and will be discussed separately in the Impact evaluation section.

Figure 1: PALS-UK logic model



Evaluation objectives

The aim of the evaluation was to assess the extent to which PALS-UK leads to an improvement in reading attainment (see Outcome measures) when comparing PALS-UK schools to business-as-usual schools. A full evaluation protocol may be found on the EEF's website (Dimova et al., 2019, 2021).

The research hypotheses this trial was able to test are as follows:

1. 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).

2. 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.

3. 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.

4. 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.

5. 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.

Owing to changes to the design outlined below (see subsequent subsection on Changes to the evaluation objectives) the impact evaluation research questions were modified. More information on the original evaluation objectives and, in turn, the evaluation design, can be found in the study protocol[3] and statistical analysis plan (SAP)[4] on the EEF website.

Changes to the evaluation objectives

In the original protocol and SAP, reading fluency and comprehension was considered a secondary outcome measure and was planned to be assessed using the WIAT-III-UK-T assessment, however this secondary outcome analysis was forgone as a result of disruption caused by the COVID-19 pandemic, which resulted in a small sample size. Due to the COVID-19 pandemic, planned outcome testing in early May-June 2020 was postponed until November 2020 – January 2021. Testing for reading fluency and reading comprehension using the WIAT-III-UK-T for those schools that were willing to participate in delayed outcome testing started in December 2020, but was not completed in January 2021 as planned, due to the movement of schools to remote learning as a result of the national lockdown from January 2021. In total, 18 schools (11 Control schools; 7 PALS-UK schools) completed the reading comprehension subtest of the WIAT-III-UK-T in December 2020. None of the schools completed the oral reading fluency subtest. Owing to the minimal statistical power arising from this small sample size and the likelihood of higher attrition following the January 2021 lockdown, the decision was taken to stop all data collection activities for assessing oral reading fluency as measured by the WIAT-III-UK-T subtests. Comprehensive description of the outcomes, including the analysis using this small sample follows in the Outcome measures and **Secondary analysis** section. Further information on these changes is captured in the updated study protocol and the updated SAP on the EEF website (see footnotes).

Ethics and trial 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.

---

[3] **https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Protocols/PALS_trial_protocol_update_2021.pdf**

[4] **https://educationendowmentfoundation.org.uk/public/files/Projects/PALS_SAP_update_2021.pdf**

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.

School recruitment was carried out by the delivery team, who at the time were based at Coventry University (Dr Emma Vardy and Dr Helen Breadmore). Parents or guardians were given the option to withdraw their children from the trial as the decision-makers for individual pupils. Prior to pupil data being sent to the delivery team, parents were sent information and withdrawal forms by the school and had the opportunity to return them. Schools sent parental information sheets and withdrawal forms to parents after signing the Memorandum of Understanding (MoU) describing what the trial involves. Parents could withdraw their children at any time from the research, but the initial withdrawal forms had to be returned by parents within two weeks. If participants chose to withdraw their children from the study later on, their data was not collected and was deleted as appropriate (see COVID-19 updated participant information sheet and withdrawal form). The recruitment documentation—including a MoU, information sheet for parents, and withdrawal form—is provided in Appendix C. The MoU and the parent information sheet were appropriately updated, with school staff and parents informed about changes to testing and data collection due to the COVID-19 pandemic.

The delivery team collected consent forms for headteachers, teachers, and other school staff who volunteered to participate in an interview. The cover page for each survey contained an informed consent and data protection statement to inform respondents that participation in the survey is entirely voluntary. Also, the surveys did not collect personal identifying information such as the respondent's name, date of birth, or contact details. Privacy notices were provided to interviewees also. Informed consent was collected at the start of each interview, and it was made clear to interviewees that participation in the case studies was entirely voluntary. Privacy notices for the surveys and interviews can be found in Appendix H: Implementation and Process Evaluation Privacy Notices.

Data sharing was governed by the data sharing agreement signed between the funder (the EEF), the delivery team (Coventry University/Nottingham Trent University), and the evaluation team (RAND Europe).

None of the members of the evaluation team had any conflicts of interest in undertaking this evaluation.

Data protection

The delivery team obtained personal data from schools and pupils at baseline as a data controller. Basic pupil information was obtained on the basis of 'public task' from schools pursuant to brief data sharing undertakings or agreements with each school that was recruited. The baseline data was then checked by the delivery team and shared with RAND Europe under a data sharing agreement with Coventry University. RAND Europe, via AlphaPlus, which oversaw test administration and data collection, obtained pupil outcome data on the basis of legitimate interests. Pupils and parents were provided with age-appropriate, fair processing privacy notices that explain the use, storage, and secure handling of the data. This also included an option to withdraw from the study.

Data was only saved on GDPR-compliant, secure servers inside the EEA or the U.K. All processes were handled in accordance with RAND's Data Protection Policy. RAND Europe is registered with the 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 the free school meal (FSM) status of pupils. RAND Europe considered 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 the EEF will become the data controller.[5] The data will be shared with the EEF's archive manager and, in an anonymised form, with the Office for National Statistics and

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

potentially other research teams. Further matching to the NPD and other administrative data including KS1 may take place during subsequent research.

Project team

The intervention was originally developed in the United States by Douglas Fuchs, Lynn Fuchs, and other researchers from Vanderbilt University. The PALS version delivered in this trial was adapted for the U.K. context by Dr Emma Vardy from Nottingham Trent University and Dr Helen Breadmore from Birmingham University who also led the PALS-UK delivery team. Both were originally at Coventry University during the early stages of this trial. The overall responsibility for recruiting and training school staff and providing support to all schools as per the terms of the sign-up information sheet ultimately rested with the PALS-UK delivery team. The delivery team was also responsible for organising and administering baseline testing. In addition, they provided advice to the evaluation team in the selection of outcome measures and on the design of the compliance measures.

The evaluation was conducted independently by RAND Europe. The evaluator was responsible for the outcome and process evaluation, trial design, analysis, reporting, and quality assurance of the evaluation. The evaluation team in RAND Europe comprised Dr Sashka Dimova (current project leader), Dr Andreas Culora (project management, field work and analysis), Lucy Gilder (field work, analysis), and Dr Alex Sutherland and Dr Sonia Ilie (both previous project leaders, formerly RAND Europe).

Methods

Trial design

The PALS-UK evaluation was designed and executed as a two-group parallel, stratified, cluster-randomised efficacy trial, with school as the unit of randomisation. In total, 89 schools were recruited from the Midlands and North East of England. The areas where recruitment took place were selected at the start of the trial.

Randomisation (in September 2019) occurred as planned and schools were assigned to either treatment (PALS-UK) or control condition. We randomised schools within regions and by school size (single-form versus multi-form entry). All schools signing up had a fifty-fifty chance of being assigned to the treatment group within each regional-form cluster (see Randomisation section).

There is only one treatment condition in this trial: throughout the 2019/2020 academic year, all Year 5 teachers working in schools allocated to the treatment condition were eligible to receive training and support to implement PALS-UK sessions in their Year 5 classroom. All Year 5 pupils were eligible for inclusion in the trial (selection criteria are outlined below, see Participant selection).

For control settings, it was business as usual. As an incentive, these schools received a payment of £500 on completion of outcome testing.

Pupils in Year 5 classes selected for the evaluation were tested twice, at pre-test (before randomisation) and post-test (in the next academic year, approximately eight months after the programme ended). All measures selected were standardised, age appropriate, and were selected by the evaluation and delivery teams (see Outcome measures).

A summary of the cRCT design can be found in Table 3.

Table 3: Trial design

Trial design, including number of arms		Two-arm, stratified and cluster-randomised
Unit of randomisation		School
Stratification variable(s) (if applicable)		Region (Midlands and North East) School size (single-form entry versus multiple-form entry)
Primary outcome	Variable	Reading attainment
	Measure (instrument, scale, source)	PiRA, 1–40 (age standardised <70–>130), designed by Rising Stars
Secondary outcome(s)	Variable(s)	1. Self-efficacy in reading 2. Reading fluency and comprehension
	Measure(s) (instrument, scale, source)	1. Reading self-efficacy self-reported questionnaire, 0–140, designed by Carroll and Fox, 2017[6] 2. WIAT-III-UK-T, reading comprehension and oral reading fluency subtest, designed by Pearson Clinical

[6] More information on the 20-item self-efficacy scale is available in Carroll and Fox, 2017.

Baseline for primary outcome	Variable	Reading attainment
	Measure (instrument, scale, source)	PiRA, 1–40 (age standardised <70–>130), designed by Rising Stars
Baseline for secondary outcome(s)	Variable	1. Self-efficacy in reading 2. Reading fluency and comprehension
	Measure (instrument, scale, source)	1. Reading self-efficacy part of a self-reported reading survey, 0–140, designed by Carroll and Fox, 2017 2. N/A

Participant selection

Schools

Schools meeting the following criteria were eligible for inclusion in the study:

- has not used PALS approach in the last three years;
- has not taken part, or will not take part, in other EEF trials for KS2 in 2019/2020 and the first half of the school year 2020/2021;
- has at least one single-form entry Year 5 class in the school;[7]
- is willing to be randomly assigned to intervention or 'business as usual' at the level of the school;
- is willing to engage with the intervention and to implement the programme;
- is willing to provide child background information to the evaluation team; and
- is willing to facilitate baseline and post-intervention data collection.

Eligible schools were accepted into the trial after they signed the MoU, which outlined the roles and responsibilities of all stakeholders involved. Schools' responsibilities and associated tasks were:

- distributing an information sheet, privacy notice, and opt-out form to parents and carers;
- collecting any opt-out forms and notifying the delivery or evaluation team immediately if a teacher or pupil wishes to withdraw from data collection; and
- accommodating and facilitating the child assessments at the beginning and end of the study.

The MoU specified that once schools agreed to participate, the expectation was that final outcome testing of children would be allowed, even if the school were to withdraw from the intervention. All schools that fulfilled the inclusion criteria and provided baseline data and completed baseline assessment were considered for randomisation.

Parent information sheets with withdrawal forms were sent to all schools in the trial by the delivery team. Schools were asked to circulate these sheets to all parents with children eligible to take part. All recruited schools confirmed prior to baseline testing that parent information sheets had been sent out.

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

The delivery team had responsibility for school recruitment. They approached around 2,000 schools in the summer of 2019. A total of 258 schools expressed an interest in the trial; of these, 98 schools returned the signed MoUs. However, nine schools did not complete the prerequisite tasks (completion of baseline testing) and were therefore not eligible for

[7] Exceptions were 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. In these instances, evaluation data was only collected for Year 5 pupils.

randomisation. In total, 89 schools completed the baseline requirements and were part of the trial for randomisation. The evaluation team then communicated the allocation result to the delivery team, which passed it on to participating schools.

Multi-entry schools

In multi-entry schools, only one class was tested and was considered to be part of this evaluation. Testing one class significantly reduced testing costs but had only a small effect on the power of the study. A class was randomly selected for testing before the baseline assessment. The same selected Year 5 class was asked to take part in testing at the beginning and end of the project. The random selection for testing was conducted in Stata.

Multi-entry intervention schools

If the school was selected to deliver PALS-UK, all Year 5 teachers at the school received the PALS-UK manual, training, resources, and support needed to deliver the intervention. Intervention schools with two or multiple-entry Year 5 classes delivered the intervention in all classes, but only the randomly selected Year 5 class was required to complete the assessment and was considered to be in the evaluation.

Multi-entry control schools

All Year 5 classes in the control schools continued implementing business as usual reading practices, but only the randomly selected Year 5 class undertook baseline and post-trial outcome testing and was included in the evaluation.

Teachers

All Year 5 teachers from intervention schools were offered the training and resources needed to deliver PALS-UK. Year 5 teachers randomised into the control group did not receive any training and instead took a 'business as usual' approach to teaching.

Children

There were no inclusion or exclusion criteria based on pupil characteristics as PALS-UK is a whole-class intervention. To minimise the burden on schools, pupils enrolled at the time of school recruitment in 2019 were included in the study, but pupils who joined the schools at a later time were not included in the evaluation as this required additional information collected from schools. This requirement was checked by comparing pupil names and UPNs at the start and end of the study to ensure that only those pupils present from the start were included in the trial. Additionally, children with severe Special Educational Needs and Disabilities (SEND) were not necessarily paired with other children but could carry out PALS-UK activities with a teaching assistant (TA). Teachers had the option to exempt these children from outcome testing.

If parents chose to withdraw their child from the study, data for that child was not collected. However, as this was a whole-class intervention, the withdrawn children still took part in the PALS-UK sessions but were not assessed. Parents had the right to withdraw their child at any time throughout the study duration.

Selection of the subsample of children that were required to take an assessment on oral reading fluency and reading comprehension

RAND Europe randomly selected ten intervention children to undertake an individually-administered assessment on reading fluency and comprehension (note that assessment on this outcome was planned but foregone, see Changes to the evaluation objectives). In multi-entry schools, this subsample of children was selected from the classroom selected for baseline outcome testing. The selection, using Stata, was such that all pupils in the classroom had equal chance of being selected irrespective of their previous attainment or other characteristics. The selection was not revealed until the time of the post-trial outcome testing (in November-December 2020): this was done to minimise the possibility that, if known too far in advance, schools might inadvertently focus resources and effort on the selected children.

Outcome measures

The selected tests were standardised, age-appropriate, and selected by the evaluation and delivery teams. The tests have an adequate level of reliability and sensitivity and were administered in their entirety, in the same order for all pupils.

The overall testing burden on pupils was up to 60 minutes for the completion of the primary outcome based on the PiRA assessment (40 minutes) and the secondary outcome based on the self-efficacy element of the reading questionnaire (15 to 20 minutes). All schools were asked to ensure that the primary outcome measure was completed first.

Outcome testing was administered by classroom teachers at two different time points. At baseline, the PiRA test was administered by Year 5 teachers in participating schools. Baseline testing took place before randomisation was completed, specifically in the first two weeks of September 2019 (between 2 and 12 September 2019) and was overseen by the delivery team. At baseline, the data was collated and scored by the delivery team.

Post-trial outcome testing was initially scheduled to take place between May and June 2020. Due to the COVID-19 pandemic, this planned outcome testing was postponed and was instead carried out in November and December 2020, once pupils were able to return to schools in September 2020. RAND Europe, in partnership with AlphaPlus, was responsible for outcome testing at endline. In the original plan, endline testing was to be administered by qualified testers from AlphaPlus that would have been blind to allocation. In light of the COVID-19 crisis, however, schools frequently adopted 'no visitor' policies; as a result, at endline, the PiRA test and the self-efficacy survey was administered by Year 6 teachers in participating schools. To ensure a higher completion rate and limit the amount of missing data, teachers were asked to facilitate one more round of testing on pupils that have missed the first round. The endline data was scored by qualified testers from AlphaPlus who were blind to allocation.

Baseline measures

Baseline assessments (for intervention Year 5 classes) consisted of the following individually administered tests:

- the PiRA test, which is sold in the U.K. by Rising Stars (RS) Assessment from Hodder Education (primary outcome measure); and
- the reading self-efficacy test (secondary outcome).

Baseline data was collected prior to randomisation.

Primary outcome measure

The primary outcome for this study is a standardised measure of reading skills of pupils based on the PiRA test.[8] The test takes around 40 minutes to complete and is designed for whole class use, with pupils of all abilities. This test assesses general reading skills (specifically comprehension, inference, language—structure and presentation). Given that the main outcome of PALS-UK is improved reading attainment (as shown in Figure 1), the PiRA is thus an appropriate primary outcome measure given the focus on general reading skills. This is a termly test which offers full coverage of the U.K. curriculum, provides standardised results and age-related scores that are benchmarked across national averages,[9] and has high internal validity and test reliability with Cronbach's alpha between 0.75 and 0.92, and has a high correlation with external measures of attainment (McCarty and Ruttle, 2018).

In this trial, the Autumn 5 version of the PiRA paper test was used at baseline while the Summer 5 version of the PiRA paper test was used at endline; this was in line with the original plan. The study team, following discussions with the EEF and delivery team, took the decision to retain the Summer 5 version—despite the fact that pupils took the test as Year 6 pupils—on the basis that it was likely that pupils' reading and literacy learning had been disrupted since March 2020 due to the shift to remote learning. Thus, the likelihood for observing ceiling effect was considered low, and Summer 5 was deemed appropriate as an outcome measure for these pupils. The PiRA 5 summer test has 30 items ranging between 0 and 1 or 0 and 2, and the total marks for the pupil's raw score is 40. Raw scores were transformed

[8] For more information see https://www.risingstars-uk.com/series/assessment
[9] For more information see https://www.risingstars-uk.com/series/assessment/rising-stars-pira-tests

to age-based standardised scores for analysis and they were standardised to a mean of 100 and standard deviation (SD) of 15. The age standardised score was considered appropriate as it takes into account the different ages, as older pupils tend to score higher, and enables meaningful comparisons with national results.

Secondary outcome measures

Two secondary outcome measures were intended to be captured as part of this study in addition to the primary outcome measure.

1. Self-efficacy in reading, using a self-reported reading survey

Given that reading confidence is a defined secondary outcome in the PALS-UK logic model (Figure 1), this measure of reading self-efficacy was deemed appropriate as a secondary outcome measure. To assess self-efficacy, a self-reported survey containing 30 items divided into two parts—reading motivation (ten items) and reading self-efficacy (20 items)—was administered to pupils. Both parts of the survey were developed by researchers from Coventry University, however, for this evaluation the delivery team made small amendments to the survey. While data on reading motivation and reading self-efficacy was collected from pupils at baseline and endline, only responses to items in the reading self-efficacy part were of interest in this evaluation. Thus, only items in the reading self-efficacy part were used to assess changes in self-efficacy in reading for the secondary analysis.

The self-efficacy part has been tested in a pilot validation study with thirty children and has an internal reliability of $\alpha = 0.912$ (Carroll and Fox, 2017).

Responses to items in the self-efficacy part of the survey could take on values between one and seven. The survey had positively and negatively worded questions. Negatively worded questions were reverse scored. Overall responses to the reading self-efficacy part of the survey could therefore take on values between zero and 140, where a higher score was equivalent to a higher level of reading self-efficacy.

To assess changes in pupil's self-efficacy, the survey was administered at the same time as the primary outcome measure. It was completed at two different time points (September 2019 and November to December 2020). The survey was compiled in machine-readable forms to allow scanning, data entry, and scoring by RAND Europe. At baseline, the survey was administered by Year 5 teachers and collated and scored by the delivery team. At endline, the survey was administered by Year 6 teachers and collated and scored by a team of subcontracted test administrators overseen by AlphaPlus.

A copy of the self-reported reading survey can be found in Appendix E.

2. Oral reading fluency and reading comprehension

Additionally, an oral reading fluency and reading comprehension test was administered to a small sample of schools in December 2020 (see Changes to the evaluation objectives).

The original design included a secondary outcome analysis of oral reading fluency and reading comprehension based on the WIAT-III-UK-T (see Burns, 2010).[10] This secondary outcome was not evaluated in this trial as outlined in Changes to the evaluation objectives section. The oral reading 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. Given that the WIAT-III-UK-T is a commercial test (sold by Pearson Education), we are unable to include the instruments in the report.

The WIAT-III-UK has been standardised on a national sample of 744 children and young people.[11] While psychometric testing of the WIAT-III-UK is thin, the U.S. version of the WIAT-III test has been extensively examined. Reliability has been assessed using the split-half reliability method, corrected by the Spearman-Brown formula (Burns, 2010). Using

[10] **https://www.pearsonclinical.co.uk/Sitedownloads/Productpdfs/wiat-iii-uk-subtests-description.pdf**
[11] **https://www.pearsonclinical.co.uk/ordering/how-to-order/qualifications/qualifications-policy.html**

this approach, subtest reliability ranges between 0.83–0.89 and 0.90–0.97, while mean reliability coefficients range between 0.91 and 0.98. Correlation coefficients for validity range between 0.60 and 0.82 (ibid.).

The evaluation team, following discussions with Pearson Education, established that WIAT-III-UK-T subtests can only be administered by professionals with a qualification code of CL2R now Qualification Level B.[11] Administration is undertaken on an individual basis and takes between 45 and 90 minutes depending on the age of the examinee and the number of subtests administered. However, given that only one subtest was administered to pupils in this trial, administration time was lower than this.

As set out in the trial protocol and SAP, to reduce the testing burden and costs, it was agreed to administer the WIAT-III-UK-T subtests to ten randomly selected children per school (for more information on the selection process see Selection of Subsample of Children). For multiple-entry schools, pupils were selected from within the randomly selected class who would complete both baseline and post-trial outcome testing.

The oral reading fluency subtest was not administered in any of the schools that took part in endline testing. Due to extra logistical challenges tied to the administration of this subtest and difficulties in accessing the required materials in time, schools were not able to participate in this subtest in December 2020. As an alternative, schools were scheduled to participate in January 2021 but were subsequently not able to due to the COVID-19 pandemic.

In December 2020, schools started completing the reading comprehension subtest of the WIAT-III-UK-T. To support a no visitor policy in schools in response to the COVID-19 pandemic, test data collection was undertaken remotely. The test was administered by qualified testers using video link with the support of school staff who were on hand to support pupils. Testers had the paper copy of the marking sheet when administering the physical test while the pupil and supporting school staff members would move with the tester through the WIAT-III-UK-T stimulus booklet, providing answers that the tester would record in the marking sheet. Tests were then scored remotely by qualified markers. In total, 18 schools were able to participate in the reading comprehension outcome testing in December 2020 (11 control schools; seven PALS-UK schools). As with the oral reading fluency subtest, the remaining schools were scheduled to participate in the reading comprehension outcome testing in January 2021 but the decision was taken to cancel this element of the outcome testing also due to the movement to home learning as a result of the COVID-19 pandemic in January 2021.

The reading comprehension subtest has nine item sets with examinees having specific start points and end points depending on their age (ages for start points and end points range from 6 to 14–17+). Items within a set were scored between 0 and 2 (sometimes 0, 1, or 2, sometimes just 0 and 2 depending on the item), with all items having a ‘don’t know’ option to differentiate incorrect answers from instances where the examinee did not know how to respond to the question. Examinees answered questions in the item set that they start at, as well as the item sets for the next two age-based item sets. However, if the sum of scores for an examinee’s start point was two or less, examinees were reversed by one start point to continue test administration. For example, if an examinee began the reading comprehension subtest at the start point for those aged ten but scored two or less in total for items in that set, they were reversed to the start point for those aged nine and administration continued from that start point. This reversal rule could be applied a maximum of three times.

Depending on the item sets administered to examinees, the maximum achievable raw scores across all ages captured by the WIAT-III-UK-T can range between 36 and 50. The maximum achievable raw scores for item sets with a starting point for those aged 10 and 11 are both 44. Raw scores were totalled and then transformed to age-based standardised scores for analysis. These scores are standardised to a mean of 100 and SD of 15. Furthermore, percentile rank (ranging from 1 to 99, median of 50) and age equivalent (providing the age at which the student’s raw score is the median score of all scores) were calculated. These metrics were transferred from the paper tests to a spreadsheet for provision to RAND Europe researchers for analysis.

Sample size

The power calculations reported in Table 4 assumed equal allocation to treatment and control groups. We also assumed an alpha of 5% and an intended 80% power to detect effects. We used two-level clustered designs and based our initial calculations on an intra-cluster correlation (ICC) of 13%. All power and minimum detectable effect size (MDES) calculations were performed using PowerUp! (Dong and Maynard, 2013). Using the parameters set in the study protocol

and with equal allocation to treatment and control, the MDES was evaluated at *d* equal to 0.193 for an ICC of 15% and 0.178 for an ICC of 11%.

The power calculations assume equal allocation to treatment and control groups. We follow the EEF guidance and assume that the trial should be powered to detect effect sizes of d = 0.20 or smaller as such studies are assigned the highest security rating by the EEF.[12] The estimation at protocol stage was based on 90 schools. With one Year 5 class per school included in the evaluation, at protocol stage we assumed an average cluster size of 28 pupil (for more information on class selection in multi-entry form school see Multi-Entry Schools subsection under Trial design). Based on previous research, we assumed the proportion of variance in Level 1 outcomes explained by Level 1 covariates R12 is 0.49 (equating to a correlation of 0.70, as per the power calculation table above) and R22 of 0.00. We used two-level clustered designs and base our calculations on an intra-cluster correlation (ICC) of 13%. Assuming a desired power of 80%, alpha of 5%, an ICC of 0.13 and a continuous, normally distributed outcome, the protocol MDES was d = 0.228. Using the parameters above and assuming that on average there are around seven FSM pupils in one Year 5 class (a rate of 25% FSM eligibility), the MDES was 0.263. As such, even though considered an efficacy trial, the study should be powered to detect meaningful differences between groups.

At randomisation, there were 89 schools and a total of 2,176 pupils with complete baseline PiRA data, that is, sufficient data to construct an age standardised baseline PiRA score for the pupil—the criterion for inclusion in the trial sample. This represents an average of 24.45 pupils per school. With the achieved sample and an ICC of 0.13, the estimated MDES is d = 0.23. As such, the MDES reported in the SAP is the same as the MDES reported in the study protocol (MDES of d = 0.23).

At analysis stage, our analytical sample was based on 66 schools and a total of 1,459 pupils. The ICC within schools based on the primary outcome measure was 0.14. This resulted in a MDES of 0.287, which is larger in size than both the MDES at protocol stage (0.228) and also larger than at randomisation stage (0.231).

As per the SAP, we conclude that the study is not powered for meaningful subgroup analysis. The MDES reported in Table 4 for the FSM subgroup is estimated for the FSM-only sample (that is, as a separate subsample). Estimating the MDES for the FSM subgroup using the same parameters but under a moderator framework (estimated in PowerUp! as the primary MDES) would yield a MDES of 0.359, which is smaller in size, but assuming moderator power calculations for a cluster-randomised trial with a binary moderator at the individual level, the between-school variance does not impact the power calculations (Spybrook, Kelcey and Dong, 2016). All MDES calculations were undertaken in PowerUp! (Dong and Maynard, 2013).

Table 4: Sample size calculations

		Protocol		**Randomisation**		**Primary analysis**	
		OVERALL	FSM	OVERALL	FSM	OVERALL	FSM
Minimum Detectable Effect Size (MDES)		0.23	0.26	0.23	0.27	0.29	0.35
Pre-test/ post-test correlations	level 1 (pupil)	0.7	0.7	0.7	0.7	0.6	0.6
	level 2 (class)	NA	NA	NA	NA	NA	NA
	level 3 (school)	0	0	0	0	0	0
Intracluster correlations (ICCs)	level 2 (class)	NA	NA	NA	NA	NA	NA
	level 3 (school)	0.13	0.13	0.13	0.13	0.14	0.14
Alpha		0.05	0.05	0.05	0.05	0.05	0.05

[12] For more information on the EEF classification of the security of evaluations, see https://d2tic4wvo1iusb.cloudfront.net/documents/evaluation/peer-review-process/Classifying_the_security_of_EEF_findings_2019.pdf

		Protocol		Randomisation		Primary analysis	
		OVERALL	FSM	OVERALL	FSM	OVERALL	FSM
Power		0.8	0.8	0.8	0.8	0.8	0.8
One-sided or two-sided?		Two	Two	Two	Two	Two	Two
Average cluster size		28	7	24.45	6.12	22.11	6.67
Number of schools	Intervention	45	45	44	44	37	34
	Control	45	45	45	45	29	27
	Total	90	90	89	89	66	61
Number of pupils	Intervention	1,260	315	1,173	278	850	220
	Control	1,260	315	1,110	267	609	147
	Total	2,520	630	2,283	545	1,459	367

Estimation of ICC

The ICC is a key parameter for clustered trials. It represents the proportion of variance in a given outcome that can be explained by the variation between clusters (here schools) as opposed to within-clusters. If an inaccurate estimate for the ICC is used, the resulting sample size estimate may be either too large or too small.

The ICCs used for the power calculations at protocol and randomisation stage reported in Table 4 are based on similar language measures used in previous EEF trials. The ICC estimation at analysis stage is based on the primary outcome measure (PiRA) and is calculated using a model similar to Equation (1) but with no predictors, accounting for the clustering of pupils in schools (the so-called 'empty model'). The observed ICC at analysis stage (0.14) is of similar size to the predicted ICC applied at the protocol and randomisation stage (0.13).

Randomisation

While the delivery team recruited schools, random allocation of schools to the treatment and control conditions was conducted by RAND researchers on 12 September 2019 and included 89 schools that had provided baseline data and an MoU signed by the headteacher/SLT. Randomisation took place after baseline testing was completed in all schools. Schools were notified of their allocation on 13 September 2019 by the delivery team.

As pre-specified in the protocol, the unit of randomisation in this trial was school with stratification according to region and form entry type (single or multi-form entry). Strata were constructed from regions (geography-based strata, Midlands versus North East) and based on school size (defined by the number of classes within a school: one form entry school versus multi-form entry schools). The stratification was undertaken to ensure that schools from the same region as well as schools with higher numbers of pupils were allocated to treatment or control evenly. Having uneven numbers of schools with one or multiple classes would mean there is a higher probability that treatment or control groups would be unequal in terms of size. To deal with unequal treatment fractions we used the command `randtreat` and the option `misfits(global)` in Stata (version 15.1; Carril, 2017).

Schools were randomly allocated to one of the two arms of the trial within each stratum. Table 5 shows actual allocations by region and school size. In total, 45 schools were allocated to the control condition of which 24 were one-form entry and 21 were multi form entry; 44 schools were allocated to the intervention (PALS-UK) condition: 22 one-form and 22 multi-form.

Table 5: Allocation of schools to treatment or control by region

		Control group			Intervention (PALS-UK) group		
		N One-form schools	N Multi-entry schools	N Total schools	N One-form schools	N Multi-entry schools	N Total schools
1	Midlands	13	18	31	12	19	31
2	North East	11	3	14	10	3	13
	Total	24	21	45	22	22	44

Statistical analysis

The outcome analysis was conducted on an intention to treat (ITT) basis (Fisher et al., 1990). The analysis included all randomised schools and 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 approach compared outcome means for the treatment and comparison groups, and subjects were 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.

Analysis was not blind to allocation; however, analysis bias was minimised by the provision of explicit Stata analysis code (and output) and the later depositing of data for potential future replication work. Full details of the analysis code and relevant output are included in Appendix G.

All analyses were performed in Stata, versions 16 onwards (Heß, 2017).

Primary analysis

The primary goal was to examine whether participation in PALS-UK produced improvements in reading attainment, as measured with the PiRA test. As specified in the SAP, age standardised scores on the PiRA were used as the outcome in the primary analysis. To estimate the impact on the primary outcome we used 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 and between schools (Hox, 1998; Snijders, 2005; Snijders and Bosker, 1994).

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

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

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 reading attainment; 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[13]; 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 in that it is a number that can take any value.

Our target parameter (that is, 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.

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

Secondary analysis

The secondary outcome analysis is based on the reading self-efficacy survey. Pupil's self-efficacy was assessed following the same specification to Equation (1) listed under primary outcome analysis above, but we substituted the self-efficacy score at endline as a secondary outcome.

The vector of pupil-level characteristics in Equation (1), X_{ij}, will include the relevant baseline measure: the baseline reading self-efficacy score.

Owing to changes outlined above, secondary outcome analysis on pupil's oral reading fluency and reading comprehension using the WIAT III subtests were not possible; that is, given the size of the sample the analysis would not have been capable of detecting relevant differences. However, we report mean outcomes for the treatment and control groups separately on the reading comprehension WIAT-III-UK-T subtest for the small number of schools that completed it in December 2020 (n = 18).

Analysis in the presence of non-compliance

The main framework of analysis for this trial is intention to treat, however, we explore the effect in schools that were allocated to the intervention group that implemented the intervention as intended, based on a compliance score. To enable a non-compliance analysis, 'compliance' was defined at the school level, based on completion of programme activities, as recorded by the delivery team.

We have defined 'compliance' as the fulfilment of a set of minimum criteria that determine whether a school has delivered PALS-UK. This is a binary measure indicating whether a school was compliant or not. These criteria, defined in collaboration with the delivery team, consist of items 1, 2, and 3 in Table 6. For a school to be deemed compliant, *all three compliance criteria* must be fulfilled. The compliance score discussed above was calculated by assigning 1 if the three 'non-negotiables' of PALS were met at the school level or 0 otherwise. Control schools were assigned a compliance score of 0 (i.e. no implementation of PALS).

The main limitation of the compliance construct is that compliance with the second and third criteria as specified in Table 6 was defined based on evidence provided only by those schools and staff that engaged in the surveys. Data on these items was collected from the delivery survey of school staff, that is, information on completion of pupil training was collected from the teacher delivery survey while information on the number of weeks PALS-UK was delivered was available from the headteacher delivery survey[14] as well as observation data supplied by the delivery team. Outside the information from the surveys, the present study did not have in place practices to record the number of weeks that PALS-UK was delivered independently from the delivery team. As a proxy measure, observation data supplied by the delivery team was used to determine the number of weeks of PALS-UK delivery in schools. However, we note that this was not a robust indicator of compliance. Based on the collected survey data, 21 schools were compliant. This is driven by low response rate on the survey (28 out of 44 headteachers responded to the delivery survey and 40 out of 44 teachers responded to the teacher delivery survey) and by the fact that not all headteachers or teachers provided a response. Missing survey data on any criteria was scored as zero.

In a situation of imperfect compliance whereby not all participating schools are deemed compliant using the three criteria, we will undertake a CACE (complier average causal effect) analysis by drawing on an instrumental variable (IV) approach, and using a two-stage least squares (2SLS) estimation approach to recover the treatment effect for those that complied with assignment. The first stage estimates if the assignment to PALS-UK pushes schools to take up treatment (the first stage regresses treatment assignment on compliance, as defined above). This provides an estimate of the compliance rate. Results for the first stage will report the correlation between the instrument and the endogenous variable as well as an F test. The second stage of the IV estimation predicts the outcome using the compliance rate estimated in the first regression by substituting the treatment indicator, PALS-UK in Equation (1), with the compliance rate (Angrist and Keueger, 1991; Angrist, 2006). The results of this model will answer the research question: to what

[14] This study collected information on the number of weeks PALS-UK was delivered from headteachers rather than Year 5 teachers. This is a deviation from the SAP.

extent does *compliance* with PALS-UK implementation requirements lead to improved outcomes for children? This model was estimated for the primary outcome measure only.

In addition to what was stated in the SAP, we also conducted additional compliance analysis to account for the high level of missing information on the third compliance ingredient, that is, completion of delivery (highlighted in red in Table 6). As a result, we specify two additional compliance indicators.

First, we carried out a compliance analysis based on a score constructed from the first two elements in Table 6, that is, schools were deemed compliant if they attended the teacher training (Criteria 1) and delivered the four weeks of pupil training (Criteria 2). We refer to this compliance score construct as 'compliance score 2' in the **Error! Reference source not found.**.

Second, we undertook a compliance analysis based on a proxy indicator using data from alternative sources—that is, using monitoring logs collected by teachers rather than data from our own survey. The present study recorded the week the peer observations were completed (highlighted green in Table 6; see Implementation and process evaluation for more information on peer observations) and consequently was considered a reasonable proxy for delivering PALS-UK for a minimum number of 12 weeks. A school was considered compliant on this element if it completed, and provided to the delivery team, a peer observation in week eight or later suggesting that a minimum of 12 weeks of PALS have been delivered when counting for the four-week pupil training. We refer to this compliance score construct as 'compliance score 3' in the **Error! Reference source not found.**.

Table 6: Compliance measure development

Number	Compliance criterion	Data source	Compliance indicator
1	Attendance at all PALS-UK initial training session	Register of attendance	Attendance at Part 1 teacher training: at least 1 member of staff per school attends.
2	Completion of the four weeks of pupil training to the manual	Teacher delivery survey	Delivery of the four weeks of training: all weeks delivered in order for school to be deemed compliant.
3	Completion of PALS-UK delivery	Headteacher delivery survey	Delivery of the main intervention: a minimum of 12 consecutive weeks delivered.
4	Completion of peer observation at week 8 or later	Observation form	Delivery of the main intervention: an observation completed at week 8 of programme delivery or later

Missing data analysis

Missing data can arise from item non-response or attrition of participants at school and pupil levels. Even though it is important to include all data, it can be problematic to apply the intention to treat principle if we are not able to complete follow up testing for all randomised schools. To deal with missing data we completed the following steps.

First, we explored the extent of missingness and then also explored the pattern of any identified missingness. The procedure outlined below refers to the primary outcome measure (and associated primary analysis); for the secondary outcome measures we reported the extent of missing data but did not undertake any additional analyses.

We attempted to follow up with all randomised schools even after having withdrawn from allocated treatment due to COVID-19 disruptions. Second, we organised mop-up testing for children that missed the initial assessment. Third, each test of the outcome measure mentioned above (PiRA, WIAT-III-UK-T oral reading fluency and reading comprehension) came with associated marking procedures that result in item non-response not being an issue (in the sense that a missing answer is considered an 'incorrect' answer and therefore does not contribute to the final score). For the reading self-efficacy measure, we addressed any item non-response by using a mean of the completed items rather than a sum score (in the absence of validation data to provide a different factor score derivation procedure).

In the baseline assessment data, we had already observed a missing data case in relation to the age-standardised PiRA score: there are pupils included in the trial who had sat the baseline assessment but obtained a raw score of 0 on the test. In consultation with the delivery team it was ascertained that these pupils were tested but did not provide sufficient answers for them to accrue any raw score other than 0 on the PiRA test. Because the lowest score that can be age-standardised using the PiRA manual is 1, these pupils therefore have missing baseline assessment data. There are 11 such observations (highlighted in the baseline equivalence table above). Our preferred approach, to avoid removing these pupils from the analysis, is to undertake what amounts to closest-value data imputation (data replacement) and assign a raw score of 1 on this test, the next possible raw score. This would allow for the calculation of an age standardised PiRA score and the inclusion of the 11 observations in the final analysis (providing outcome data is forthcoming for these pupils).

Given that the outcome data was incomplete (see Sample size section for more information), we first determined the proportion of missing data in the trial. We explored attrition across trial arms as a basic step to assess bias (Higgins et al., 2011). We also provided cross-tabulations of the proportions of missing values on all outcome measures.

To assess whether there were systematic differences between those who did not provide a valid PiRA post-test score and those who did—and thus whether these factors should be included in analysis—we modelled missingness through a logistic regression model at follow-up as a function of baseline covariates, including treatment. The analysis model for this approach mirrored the multilevel model in the primary outcome analysis (pupils clustered in classes), but the outcome was a binary variable identifying missingness (yes/no). This also made use of the fact that baseline data relevant to the primary outcome analysis was complete for all trial participants. If any of the baseline covariates were seen to be statistically significantly associated with missing primary outcome data (at the 5% level), then the primary analysis was repeated adjusting additionally for these covariates as fixed effects.

If there was less than 5% missingness overall (that is, the primary analysis model included at least 95% of randomised pupils), we would carry out a complete-case analysis and undertake an exploratory robustness analysis using a full-information maximum likelihood (FIML) approach (instead of multiple imputation, MI) because FIML can be estimated in a single model and simulation studies (Allison, 2003) show that it can reduce bias as well as MI.[15]

If there was more than 5% missingness overall, we proposed to undertake analysis to understand if the data appears to be missing completely at random (MCAR) or whether the weaker missing at random (MAR) assumption applies. We assessed if the data was consistent with being MCAR by using the `mcartest` Stata package; this implements Little's test of MCAR. We would use this as an indication only, as opposed to a definitive conclusion, and would supplement this by creating dummy variables to identify missingness for the primary outcome variable (Allison, 2001) and then explore (through simple t-tests) if this missingness indicator was associated with any of the other variables in the dataset, adjusting for multiple comparisons. This would allow us to understand if the missing data pattern was MCAR or if the main outcome was missing conditional on other variables.

If the missing data pattern appeared to be MCAR, we would have estimated Equation (1) as stated, without any further accounting for the missing data, as the results will have been unbiased.

It is difficult to show in practice if data is MAR or missing not at random (MNAR) given the very data that is missing is needed to ascertain this (Fielding et al., 2006). Therefore, if missing data was not MCAR (based on the tests above and could be either MAR or MNAR) we proposed to run a pattern mixture model. This approach models for the observed and unobserved portion of the missing data (Carpenter and Kenward, 2012) and can be undertaken in Stata by first using the `rctmiss` (White, 2018) package which can model data and missingness jointly with a pattern mixture model whereby the differences between the missing and observed data are modelled at the same time as the main effect is estimated.

[15] Multiple imputation is not necessarily the gold standard in missing data handling in RCTs, with other (simpler) methods providing similarly unbiased estimates (Sullivan et al., 2018).

Subgroup analyses

As defined in the trial SAP we also conducted the analysis for the following subgroups, using the same model as our primary analysis.

1. **Children who are eligible for FSM** as registered in the National Pupil Database (NPD). We explored differential effects for FSM pupils as they are considered a key target group by the EEF using the variable EVERFSM_6.

2. **Children who are registered as EAL in the NPD**. The EAL indicator identifies children who are routinely exposed to other languages in their home. Evidence suggests (Strand and Hessel, 2018) that it is not a good indicator of pupils' later attainment because of the heterogeneous group it captures, with English proficiency emerging as a better predictor of later attainment instead. However, the focus on EAL is important as an earlier evaluation of a PALS programme in ten schools in the U.S., which focused on English Language Learners (ELL) pupils with learning difficulties, found a large positive effect for this subgroup (Sáenz, Fuchs and Fuchs, 2005).

3. **Children with lower and higher reading ability who score below and above a threshold point on the PiRA age-standardised scores during the baseline test**. Finally, we explored if programme effects differ across high and low reading achievers as previous findings about the effects of pairing low- with high-ability pupils are mixed (Lou et al., 1986). The evaluation of PALS-UK informs if there are any differences in programme effect for lower reading ability pupils compared to higher ability pupils. In the analysis, pupil ability was based on the PiRA reading assessment scores at baseline, as outlined above. Low reading achievers are pupils scoring below average on the age standardised PiRA test, while high reading achievers are pupils scoring above average. We constructed a binary variable (below the threshold; at or above the threshold) to enter into an analysis. These binary variables were constructed based on the sample-specific mean PiRA score to ensure as much as possible that even samples could be achieved for the analysis. Additionally, we undertook an exploratory analysis (see Additional analyses below) for the very low ability readers (bottom quartile of PiRA scores).

As mentioned in the SAP, all subgroup analyses in this evaluation were exploratory and likely underpowered.

Analytical approach

As a basic descriptive step, the study reports mean outcomes by subcategories of children eligible for FSM, children with EAL, and low reading achievers at baseline.

To estimate differential programme effects for the subgroups of interest we added an interaction term in the main equation (1) above for all possible values taken on by the treatment and the subgroup variable of interest. For example, for treatment indicator and FSM eligibility status, we include the product of these two variables to produce a model that has one covariate for every possible group—that is, four groups: (1) pupils eligible for FSM in PALS-UK schools, (2) pupils eligible for FSM in control schools, (3) non-FSM pupils in PALS-UK schools, and (4) non-FSM pupils in control schools. FSM eligible pupils in control schools were specified as reference group in our model (i. e. $PALS_0 * EVERFSM\_6_1$). The subgroup model specification is as follows:

(2) $Y_{ij} = \beta_0 + Z_j\beta_1 + X_{ij}\beta_2 + PALS_1 * EVERFSM\_6_1\gamma_2 + PALS_1 * EVERFSM\_6_0\gamma_3 + PALS_0 * EVERFSM\_6_0\gamma_4 + X_{ij}\beta_6 + u_j + e_{ij}$

For the reference group in our model, β_0 is the intercept, $\beta_1$is the slope for school effects, β_2 is the slope for previous attainment, and γ_2 identifies the observed differential effect on reading attainment for pupils eligible for FSM in the PALS-UK group $(PALS_1 * EVERFSM_{6_1})$ in comparison to pupils eligible for FSM in the control group. A similar model was estimated for each of the subgroups (EAL and reading ability). To obtain the effect size, the difference in means between FSM eligible pupils in the PALS-UK schools and those in control schools was divided by the pooled variance in the estimation sample.

We also undertook the subgroup analysis by splitting the sample and using the sample of interest only (i.e. as a separate subsample). We estimate programme impact on the primary outcome for the subgroups above using separate models.

These subgroup analyses are undertaken for the main outcome measure only. The statistical analysis plan acknowledges that they would most likely be underpowered and, as a result, are exploratory.

Additional analyses

We carried out the following additional exploratory analyses.

Subgroup analysis

As an additional analysis, we estimate the impact of PALS-UK on pupils with very low reading ability at baseline, defined as the bottom quartile on the baseline PiRA score. The estimation approach mirrors the approach for the subgroup analysis, substituting the subgrouping variable in the interaction term with treatment with a very low ability variable, taking the value 1 if the pupils scores below the 25th percentile on the baseline PiRA test and 0 otherwise.

Compliance analysis

In addition to the original compliance analysis as specified above, two different model specifications were used to estimate the effect of PALS-UK on the primary outcome measure for schools deemed compliant with programme delivery. We estimate the effect of PALS-UK on the primary outcome measure for schools that are compliant based on the first two elements in Table 6, that is, schools were deemed compliant if they attended the teacher training (Criteria 1) and delivered the four weeks of pupil training (Criteria 2). Additionally, given the high missingness in the surveys, we estimate the effect of PALS-UK on the primary outcome measure in schools deemed compliant using information from the peer observations instead, that is, the completion of peer observation at week eight or later (Criteria 4). See Analysis in the presence of non-compliance for more information.

Estimation of effect sizes

We used the effect size for cluster-randomised trials given in the EEF statistical analysis guidance for evaluators' (EEF 2018b):

$$s^* = \frac{(\bar{Y}_T - \bar{Y}_C)_{adjusted}}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}}$$

Where s^* represents the clustered effect size, $(\bar{Y}_T - \bar{Y}_C)_{adjusted}$ is the mean difference between intervention groups adjusted for baseline characteristics (that is, the coefficient from the multilevel model), and the denominator is an estimate of the pooled variance. A 95% confidence interval for the ES that takes into account the clustering of pupils in schools is reported.

Implementation and process evaluation

The logic model illustrated in the Evaluation overview

The evaluation was designed and executed as a stratified, two-arm, cluster-randomised controlled trial (cRCT), across 89 schools from two geographical regions (the Midlands and the North East). Of the 89 schools recruited to PALS-UK, the programme was delivered in 44 intervention schools with the remaining 45 schools assigned to a business as usual (BAU) control condition. In this evaluation, PALS-UK was delivered to Year 5 children. This evaluation of PALS-UK was comprised of three strands: (1) an impact evaluation, to assess whether PALS-UK led to an observable improvement in outcomes for Year 5 pupils, (2) an implementation and process evaluation (IPE) which provided contextual information around programme delivery and the 'business as usual' scenario during the trial, and (3) a cost evaluation, which provided information on the delivery costs incurred by intervention schools due to delivering PALS-UK. The movement to remote learning as a result of the COVID-19 pandemic in March to July 2020 and in January 2021 effected the outcome data collection activities and consequently the extent to which we can be confident in our findings. As a

consequence, analysis on one of the secondary outcomes—the Wechsler Individual Achievement Test–III U.K. for Teachers (WIAT-III-UK-T)—was not feasible. Further details around the trial design changes can be found in the updated evaluation protocol (Dimova et al., 2021).

The evaluation was funded by the Education Endowment Foundation (EEF). The programme was delivered by the PALS-UK delivery team (Nottingham Trent University and Coventry University) and the trial was carried out by RAND Europe; the independent evaluator appointed by the EEF.

Intervention section (Figure 1) identified the core components of the intervention, key principles against which to measure fidelity, expected moderators and mediators, and the linkages between these elements. The aim of the IPE was to address each of these elements through a range of methodologies and drawing on varied data sources. The purpose of the process evaluation was to address the following questions:

RQ 1: Was the intervention implemented with fidelity in the treatment condition 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?

IPE methods overview

To address the IPE research questions outlined above, we have developed a multi-stage mixed-methods approach that drew upon a combination of observations, surveys, case study interviews, and a review of resources and data collected by the delivery team.

In light of the delayed testing due to the COVID-19 outbreak it was agreed by all parties that an additional round of IPE data collection was needed to assess and better understand how schools worked between the end of the trial and outcome testing. The main objective of the additional IPE data collection activities is to assess and understand the activities that schools undertook since the end of PALS-UK delivery and the start of the social distancing rules, to further contextualise the findings from the trial. Further details of the original IPE activities and changes made can be found in the Evaluation Protocol and Evaluation Protocol (amended) available on the EEF website.[16]

Table 7 presents an overview of the rescoped IPE data collection activities alongside the specific aims of these activities. We have indicated in the table which activities were completed before the trial (pre-delivery phase); during or immediately after the trial (delivery phase), and in the period before outcome data collection (COVID-19 driven phase). Further details of the data collection methods are discussed in further detail below.

Table 7. IPE methods overview

Research methods	Data collection methods	Participants and data sources	Data analysis methods	IPE research questions addressed	Implementation/logic model relevance
Pre-delivery phase					
Surveys (pre-delivery, round 1)	Online survey	Headteachers	Quantitative descriptive analysis, including frequency counts, proportions.	RQ4	Documenting any interventions that may have impact on reading attainment.
Observations of two initial	Observation	Teachers	Thematic analysis.	RQ1; RQ2	Documenting quality of training and teacher engagement.

[16] **https://educationendowmentfoundation.org.U.K./projects-and-evaluation/projects/PALS-UK/**.

PALS-UK training sessions					
PALS-UK manual and resources	Document review	Delivery team	Thematic analysis.	RQ2	Quality of PALS-UK resources.
Delivery phase					
Observation of one top up training session	Observation	Teachers	Thematic analysis.	RQ1; RQ2	Documenting quality of training and teacher engagement.
Surveys (post-delivery, round 2)	Online survey	Headteachers (T)	Quantitative descriptive analysis, including frequency counts, proportions, measures of central tendency.	RQ2	Exploring experience of schools in delivery; reporting on costs associated with delivering PALS-UK.
		Headteacher (C)		RQ4	Documenting counterfactual.
		Teachers (T)		RQ1; RQ2; RQ3	Exploring experience of schools in delivery including compliance, fidelity, and impact.
Case studies (round 1)	Interviews	Headteachers (T)	Thematic analysis; within-case analysis; cross-case analysis.	RQ1; RQ2; RQ3	Exploring experience of schools in delivery; reporting on costs associated with delivering PALS-UK.
		Teachers or peer observers (T)		RQ1; RQ2; RQ3	Document the quality of the intervention including compliance and fidelity and exploring perceptions on impact and drivers of change.
Review of data collected by the delivery team: RA and peer observation logs; pupil pairing logs; gap task logs.	RA and peer observation fidelity score; logs	RA and peer observers	Quantitative descriptive analysis for case study schools	RQ1; RQ2; RQ3	Dosage and compliance; quality of intervention.
COVID-19 driven phase					
Surveys resulting from COVID-19 trial design changes (post-delivery, round 3)	Online survey	Headteachers (T)	Quantitative descriptive analysis, including frequency counts, proportions, measures of central tendency.	RQ3	Documenting school's experience with literacy/reading support since the start of social distancing rules.
		Headteachers (C)		RQ4	Documenting school's experience with literacy/reading support since the start of social distancing rules.
		Teachers (T)		RQ3	Exploring experience of Year 6 classes with reading support in COVID affected environment; exploring perceived impact of PALS-UK.
Case studies resulting from COVID-19 trial design changes (post-delivery, round 2)	Interviews	Headteachers (T)	Thematic analysis; within-case analysis; cross-case analysis	RQ1; RQ2; RQ3	Exploring experience of schools in delivery; reporting on costs associated with delivering PALS-UK.
		Teachers or peer observers (T)		RQ1; RQ2; RQ3	Document the quality of the intervention including compliance and fidelity and exploring perceptions on impact and drivers of change.

Review of PALS-UK manual and resources

The intervention manual was finalised and shared with RAND Europe before the training begun. RAND reviewed the manual to build better understanding of the programme and its key components, which informed later IPE tasks, such as the training observations and the phrasing of wording of survey and interview questions.

Observation of training of Year 5 teachers

From each school, Year 5 teachers and one peer observer—who could be any nominated teacher from the school—attended initial training and top-up training sessions. Four training sessions were arranged with around twenty schools and forty teachers attending each training session (although this varied between sessions). All Year 5 teachers plus peer observers were given the option to attend. RAND Europe observed the training (see Table 7) to gather information about delivery and teacher engagement as the latter may have affected the delivery of sessions and, consequently, the students' outcomes. This was recorded in the form of contemporaneous, structured notetaking rather than video recording.

Online surveys with school staff—headteachers, teachers, and peer observers

The surveys collected data on usual practices, attitudes, perceptions, and reading-related activities in the classroom from headteachers and teachers. Surveys were developed by the project team at RAND Europe in consultation with the delivery team and the EEF. The questions were tailored to each type of respondent in each group (intervention and control). Surveys were kept as short as possible and it took no more than 15 minutes to complete each. All surveys were uploaded onto the SmartSurvey platform by RAND Europe and completed online. Survey data was collected at three different time points. Headteachers from intervention and control schools were surveyed at baseline in October 2019 (baseline survey), post-intervention from June to August 2020 (delivery survey), and in November 2020 to March 2021 (follow-up survey). In addition to headteachers, teachers were surveyed in intervention schools from June to August 2020 (delivery survey) and again in the period November 2020 to March 2021 (follow-up survey). For an overview of how many headteachers and teachers participated in the surveys see Table 8. It should be noted that response rates in the delivery survey and, to an even greater extent in the follow-up survey, were lower, likely due to disruptions related to the COVID-19 pandemic.

Table 8: Number of responses and achieved response rate by survey round

	Baseline survey		Delivery survey		Follow-up survey	
	Number of respondent	Achieved response rate	Number of respondent	Achieved response rate	Number of respondent	Achieved response rate
Headteacher treatment	36	81.81%	28	63.63%	18	40.9%
Headteacher control	37	82.22%	32	71.11%	8	17.78%
Teacher treatment	Not applicable	Not applicable	40	90.09%	21	47.72%

Baseline surveys

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 headteachers at baseline was collected in October to November 2019. This was scheduled after randomisation and before programme delivery.

Post-delivery surveys

The second headteacher survey helped us to determine if control schools have changed their literacy practices in the last school year. We 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 allowed us to corroborate information obtained from teacher surveys. The survey allowed us to understand the potential costs associated with the intervention through asking about additional support time and costs (see Costs). The survey was administered in May 2020 and in total 32 headteachers took part, although ten of these respondents indicated that they had a different role in the school, for example, member of the SLT or teacher, so the sample size for most survey responses is 22.

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

- **perception of training**—how useful they believe the training has been in preparing them to deliver the intervention;

- **perception of programme effect**—the effect of PALS-UK on peer relationships and whether the strength of these relationships was affected by the intervention; it asked questions about the usefulness of each element of the programme: partner reading, re-tell, paragraph shrinking, and prediction relay;

- **use of online resources and ongoing support—**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 (for example, videos and gap tasks)—**to determine if amendments to the resources provided are needed in the future;

- **support from research assistants to Year 5 teachers**—the survey asked Year 5 teachers if support from RAs has been available; and

- **post-trial implementation**—whether schools intended to continue implementing PALS-UK, or elements of it, in the future.

Follow-up surveys

This round of survey data collection was similar to the original IPE post-delivery round and comprised:

- a survey of headteachers in control and intervention schools asking about general reading and literacy work since the start of the lockdown and social distancing; and

- 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, and catch-up and also their experiences working with children since their return to school.

Case study schools

Selection

A small number of intervention schools were studied in greater depth as purposive case studies to give further depth of information on the implementation of PALS-UK. Schools were selected randomly for case studies from a list of schools with varying leadership engagement and enthusiasm for the project as preliminarily reported by the delivery team based on interactions and communications with schools.. We also considered region and school size in our case study sampling framework in order to allow schools from different regions to be studied in more depth.

Semi-structured in-depth interviews

School staff from case study schools—headteachers, teachers, and peer observers—were invited for semi-structured in-depth interviews to gain a detailed understanding of their experiences and perceptions of the trial. Interviews lasted between 20 and 35 minutes and took the form of a guided conversation using a semi-structured interview guide. With the participants' consent, interviews were audio-recorded for the purpose of writing up accurate notes. Interviews were analysed thematically according to the questions explored by the researcher conducting the interview. The interviews helped determine the significance of the drivers of change as identified in the logic model.

We undertook two rounds of interviews. A second round was added after testing was delayed with the aim to understand in more depth any additional support related to reading in the period between the end of the intervention and the outcome testing, to further contextualise the findings from the trial. As specified in the updated study protocol, we were planning to follow the same schools in the second round, however, only one of the schools accepted the offer to take part in follow-up interviews. As a result, we invited schools that were selected on the waiting list in the first round. For an overview of the number of headteachers, and teachers interviewed by phase see Table 9. As can be seen from below, we undertook case studies in three schools between May and June 2020 and six case studies in December 2020. Only one school was followed two times.

To provide clarity on the source of statements and expressed perspectives, throughout the report, interview evidence is referenced in the form of 'TS#-1', 'or 'HTS#-2' where the first initials indicate the role of the individual (teacher or headteacher) while S# identifies case study school (S1 is School 1). The last number identifies if the interview was undertaken immediately after delivery (-1) or in the COVID-19 phase (-2). Individuals interviewed have been anonymised in relation to specific points discussed and views expressed for the purposes of respecting informed consent.

Table 9: Number of interviews undertaken with case study schools

	Post-delivery phase	COVID-19 phase
Number of headteachers	2	2
Number of teachers	4	5
Number of schools	3	6

Review of research assistant and peer classroom observation

Classroom observations were completed in all schools allocated to the intervention and were organised by the delivery team. 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 ran training on how to do the observation session. RAs completed one observation in the first four weeks and another in the following four weeks. Peer teacher observations were scheduled for around week seven and twelve of programme delivery. The observation log recording the number of observations completed and the week the observation took place was used in the construct of the compliance measure (see Analysis in the presence of non-compliance). Additionally, we have explored differences in the checklist score for the case study schools.

Pupil pairing logs

While defined as an IPE data collection method in the trial protocol, it was decided against collecting pupil pairing logs in a robust way because it would have added an additional burden on schools. The decision was made after it was assessed that it would have been challenging for schools to collect, store, and share the data in a GDPR compliant manner. However, the evaluation team were able to gather insights around pupil pairing from the endline survey of teachers and headteachers in intervention schools and case study interviews conducted with teachers in intervention schools.

IPE methods 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 were aggregated and summarised, with results used to provide a description of how the intervention worked in practice. Additionally, any open-text responses were analysed using a general inductive approach, with results used to provide a description of how the intervention worked in practice (Thomas, 2006). In addition, the IPE data was 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 allows 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; Humphrey et al., 2016, p.1).

Costs

Average marginal costs per pupil per year were estimated in two stages: first, from a discussion and data provided by the delivery team and then, second, using data collected in the IPE surveys from school staff relating to extra direct costs or additional hours worked. Average marginal costs per pupil per year over a three-year period were estimated in line with the convention followed by other EEF projects. Estimated costs are adjusted by the year when they are incurred to account for changes in value-of-money over time and inflation (EEF, 2019). The EEF cost rating is included in Appendix A: EEF cost rating. Headteachers were also asked if other school staff, such as teachers, had been required to work longer hours. These estimates together provided a general indication of the costs involved with implementing

PALS-UK. We recognise that in the absence of complete data on costs from all schools (not all schools responded to the cost-querying survey), costs may differ between schools—particularly between one-form-entry and multi-form-entry schools.

The following cost categories were explored using questions to school staff:

- direct financial costs, including direct costs paid for training and materials if the intervention were to be offered without subsidy; and
- assessment on any pre-requisite costs (such as training costs and materials) and any direct and marginal costs that are directly attributable to schools' participation;
- direct time costs, including staff time.

We have used this information to estimate cost per-pupil, following EEF guidelines on cost evaluations (EEF, 2019).

Timeline

The trial timeline is illustrated in Table 10.

Table 10: Trial 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 and 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
March–September 2020	Schools move to remote learning due to the COVID-19 pandemic	N/A
May–June 2020	Interviews in case study schools (Year 5 teachers, peer observers, headteachers)	RAND Europe
June–August 2020	Collection of Delivery surveys from Year 5 teachers 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
November 2020–March 2021	Collection of Follow-up COVID-19 surveys from Year 5 teachers and headteachers	RAND Europe
January–March 2021	Schools move to remote learning due to the COVID-19 pandemic	N/A
May 2021	Submission of draft EEF Report	RAND Europe
December 2021	Final EEF Report	RAND Europe

Impact evaluation

Participant flow including losses and exclusions

The participant flow diagram is presented in Table 11. The diagram shows that 89 schools were randomised in total. Of these, 44 (1,173 children) were assigned to the intervention group while 45 (1,110 children) were assigned to the control group. Thus, across these 89 schools, a total of 2,283 pupils were randomised in this PALS-UK trial. However, 108 of these 2,283 pupils that were randomised did not have baseline PiRA data due to being absent on the day(s) of baseline testing (98 out of 108), while a few pupils had scores that could not be standardised (6 of 108) and the remaining pupils were not administered baseline assessment due to the pupils' Special Educational Needs (4 of 108). This left an effective sample of 2,175 randomised pupils that had complete PiRA data at baseline. This difference in sample size has implications for the **Missing data analysis**, which is discussed further down in this section.

Limited numbers of schools agreed to participate in the delayed outcome testing as a result of the COVID-19 pandemic. At the time of analysis, primary outcome data had been collected from 66 schools and 1,459 pupils. The reduced sample size of the final analytical sample was due to high levels of attrition owing to the COVID challenges that schools faced.

Attrition

In total, 23 schools did not participate at outcome testing. These schools are not included in the final analysis and are listed under the 'lost to follow-up' category in the participant flow diagram.

Baseline data was available for 1,173 pupils in 44 intervention schools and 1,110 pupils in 45 control schools. At endline, data was available for 850 pupils in 37 intervention schools and 609 pupils in 29 control schools (Table 12).

The overall ratio between pupils analysed and pupils randomised is 1,459 to 2,283—a retention rate of 63.9% and an attrition rate of 36.1%. In the control group, the retention rate was 54.9% (an attrition rate of 45.1%) with 609 pupils in the analysis from the 1,110 randomised. In the intervention group, the retention rate was 72.5% (an attrition rate of 27.5%) with 850 pupils analysed of the 1,173 randomised to the intervention condition. Thus, at both the pupil and school level, the attrition rate was substantially higher in the control group as compared to the PALS-UK group.

The main reasons children were not tested were:

- 23 schools withdrew and did not agree to take part in the delayed outcome testing (25.84% of schools); or
- pupils were absent at either the baseline or endline testing, or both.

As shown in Table 12, a total of 824 pupils were not included in the final analytical sample due to attrition. Of these, 72.0% (593) were lost due to school-level attrition from the trial. Given that the school-level attrition was higher in the control group, pupils in schools in the control group naturally comprised the majority here (411 of these 593 pupils). The remaining 28.0% (231) were not included in the final analytical sample due to pupil-level attrition—absence from baseline or endline PiRA testing, or both. The most common mechanism for pupil-level attrition was absence only at endline testing (169 of these 231 pupils had complete baseline PiRA data but were absent from the endline PiRA testing). After this, it was slightly more common for pupils to have missing baseline PiRA data but complete endline PiRA data (44 out of 231), while a minority had missing PiRA data at baseline and endline (18 out of 231). Pupil-level attrition appeared to be higher among pupils in PALS-UK schools (141 of these 231 pupils). Nonetheless, the key messages here are that the main mechanism of missingness in the final analytical sample was school-level attrition, which accounted for almost three-quarters of the missing data between randomisation and analysis, and this school-level attrition was higher among pupils schools in schools in the control group. The likely explanation for school-level attrition, and at least some part of the pupil-level attrition caused by absence during the endline testing, was the challenges schools faced as a result of the COVID-19 pandemic, for example, the shift to remote learning, staff and pupil illness, self-isolation, home schooling, and so forth.

Table 11: Participant flow diagram

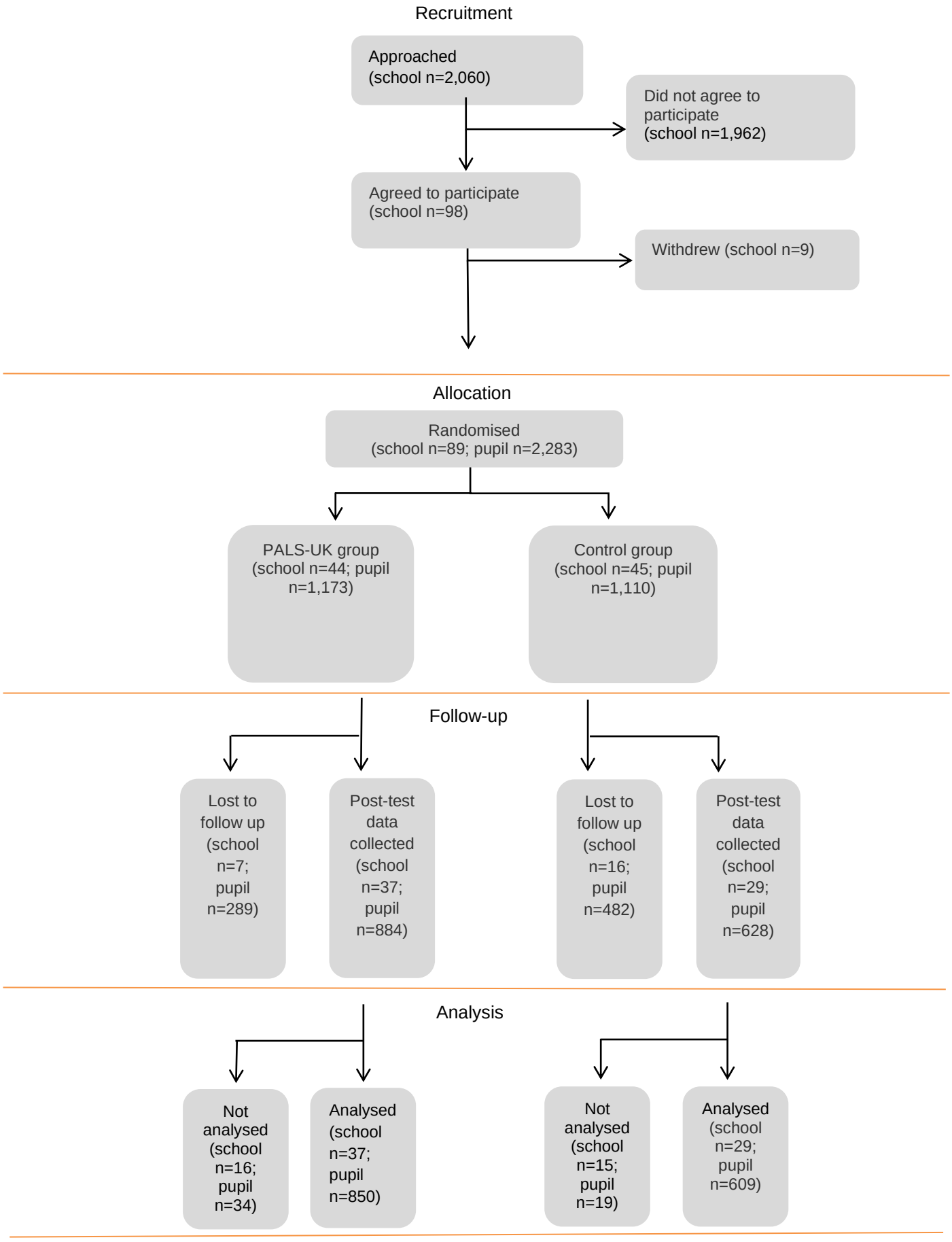


Table 12: Pupil- and school-level attrition from the trial (primary outcome)

		Intervention	Control	Total
Number of pupils	Randomised	1,173	1,110	2,283
	Analysed	850	609	1,459
Pupil attrition (from randomisation to analysis)	Number	323	501	824
	Percentage	27.5%	45.1%	36.1%
Number of schools	Randomised	44	45	89
	Analysed	37	29	66
School attrition (from randomisation to analysis)	Number	7	16	23
	Percentage	15.9%	35.6%	25.8%

Pupil and school characteristics

Table 13 shows the baseline distribution of school and pupil characteristics across the control and intervention schools using pre-randomisation data available on each respective variable.

To assess the balance of pupil characteristics at baseline in accordance with EEF and CONSORT[17] guidelines we provide descriptive tables of the distribution of pupil characteristics in the control and treatment schools. Given appropriate randomisation procedures were followed, any differences between control and treatment groups at baseline will be, by definition, due to chance, and classical statistical testing is therefore unnecessary, and imbalance may be gauged from the differences in means. Instead, tables of the means (and standard deviation, where appropriate) for each characteristic are presented. We also explore the magnitude of any differences in pupil-level pre-test measures as effect sizes (see Senn, 1994).

Balance was observed at the school level among the groups as randomised in terms of proportion of pupils speaking English as an additional language (EAL) and eligible for free school meals (FSM), with differences across intervention and control schools at 0.4 percentage points or below across these characteristics: for EAL, the difference was 0.4 of a percentage point, 16.6% and 16.2% respectively; for FSM it was 24.9% and 25.1% respectively. Similarly, balance at baseline was observed in terms of the Year 5 proportion of FSM and EAL pupils in intervention and control schools (see Appendix D: Baseline equivalence).

Baseline equivalence at the school level was also assessed along the lines of KS2 reading attainment as measured by the average scaled score. Balance was observed here also, with a difference of just one scaled score observed between intervention and control schools (104.6 and 103.6 respectively). The distributions by treatment condition for KS2 reading attainment at the school level can be found in Appendix D: Baseline equivalence.

At the pupil level, baseline equivalence was assessed in relation to gender, age (in months), baseline PiRA scores, and baseline self-efficacy survey scores. Looking first at gender, Table 13 shows a 2.0 percentage point difference in the proportion of males among intervention and control schools (52.0% and 48.0% respectively). Balance at baseline was also observed in relation to FSM and EAL status.

At the pupil level, moderate imbalance was observed along the lines of age, with a mean age (in months) among pupils in intervention schools of 113.4, while this was 114.0 for pupils in control schools. However, balance was observed across intervention (mean PiRA score of 102.7; mean self-efficacy score of 102.6) and control schools (mean PiRA score of 103.0; mean self-efficacy score of 102.6) in terms of PiRA age standardised score (ES = 0.02) and self-efficacy

[17] **http://www.consort-statement.org/checklists/view/32-consort/510-baseline-data**

survey scores (ES = 0.00). The distributions by treatment condition for each of these three pupil-level characteristics can be found in Appendix D: Baseline equivalence.

Table 13: Baseline characteristics of groups as randomised

School-level (continuous)	National-level mean	Intervention group		Control group		
		n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)	
School proportion English as an additional language (EAL)	16.0%	44/44	16.6 (24.6)	43/45 (2)	16.2 (23.4)	
School proportion eligible for free school meals (FSM)	25.0%	44/44	24.9 (16.2)	43/45 (2)	25.1 (16.1)	
Year 5 proportion EAL in the school	NA	44/44	14.7 (24.0)	45/45	16.3 (24.2)	
Year 5 proportion FSM in the school	NA	44/44	26.1 (19.7)	45/45	27.6 (19.4)	
KS2 reading average scaled score	104.4	42/44 (2)	104.6 (3.1)	41/45 (4)	103.6 (3.0)	
Pupil-level (categorical)		n/N (missing)	Count (%)	n/N (missing)	Count (%)	
English as an additional language status						
Non EAL pupil	84.0	920/1,106 (67)	83.2	882/1,057 (53)	83.4	
EAL pupil	16.0	186/1,106 (67)	16.8	175/1,057 (53)	16.6	
Free school meals status						
Non FSM pupil	75.0	825/1,107 (66)	74.5	766/1,057 (53)	72.5	
FSM pupil	25.0	282/1,107 (66)	25.5	291/1,057 (53)	27.5	
Gender						
Male	52.6	541/1,049 (124)	51.6	498/1,033 (77)	48.2	
Female	47.4	508/1,049 (124)	48.4	535/1,033 (77)	51.8	
Pupil level (continuous)		n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)	Effect size [95% CI]
Age (in months)	NA	1,113/1,173 (60)	113.5 (3.7)	1,062/1,110 (48)	114.0 (3.6)	NA
PiRA age standardised score	NA	1,113/1,173 (60)	102.7 (14.0)	1,062/1,110 (48)	103.0 (14.0)	0.02 [-0.06 – 0.10]
Self-efficacy survey score	NA	1,059/1,173 (114)	102.6 (23.0)	1,012/1,110 (98)	102.6 (22.6)	0.00 [-0.09 – 0.08]

As set out in the EEF guidelines, we also present the baseline distribution of school and pupil characteristics across the control and intervention schools available on each respective variable using the sample of schools and pupils as analysed. Table 14 presents these results.

Balance was observed at the school level among the groups as analysed in terms of proportion of pupils eligible for FSM, with differences across intervention and control schools at 1.5 percentage points (24.7% and 23.2% respectively). However, there was a greater difference observed at baseline across the intervention and control groups when looking at proportion of pupils using EAL; here, the difference was 7.2pp, with the average proportion higher in the intervention group. The school level percentage of FSM and EAL pupils by treatment condition as analysed are presented in Appendix D**:** Baseline equivalence.

Similarly, balance at baseline was observed in terms of the Year 5 proportion of FSM pupils across the intervention and control condition—27.2% and 25.6% respectively—representing less than two percentage points in difference. The difference in proportions of EAL pupils was larger. Indeed, the mean and standard deviation proportion was considerably higher among schools in the intervention group (M = 16.0%, SD = 25.9%) as compared to schools in the control (M = 11.5%, SD = 16.5%). However, this difference of 5.5 percentage points is not necessarily large enough to consider the final analytical sample imbalanced along these lines.

Baseline equivalence at the school level among the sample of 66 schools included in the endline analysis was also assessed in relation to KS2 reading attainment as measured by the average scaled score. Balance was observed here also, with a difference of just 0.5 of a scaled score observed between intervention and control groups (104.2 and 103.7 respectively).

At the pupil level, baseline equivalence was assessed for gender, age (in months), baseline PiRA scores, and baseline self-efficacy survey scores among the sample of 1,459 that were included in the primary outcome analysis. Looking first at gender, Table 14 shows a 5.1 percentage point difference in the share of males among the intervention and control groups (50.7% and 45.6% respectively). The same percentage point difference can be observed for share of females, albeit in the opposite direction (49.3% and 54.4% respectively).

In terms of FSM balance at baseline, the proportion of FSM pupils was slightly higher in the intervention group (25.9%) as compared to the control condition (24.3%), but by less than two percentage points. For EAL pupils, however, in the final analytical sample the proportion was considerably higher in the intervention group (18.4%) as compared to the control group (11.9%), representing a difference of 6.5 percentage points.

At the pupil level, imbalance was observed in relation to age (ES = 0.16), with a mean age (in months) among pupils in the intervention group of 113.5 while this was 114.1 for pupils in control schools. Balance was observed across the intervention and control groups in terms of PiRA age standardised score (ES = 0.06). The mean score in the intervention group was 103.0 while the mean score in the control group was 103.8. Finally, balance was observed among the self-efficacy survey scores (ES = 0.04) with the intervention group having a mean score of 102.0 and the control group having a slightly higher mean score of 102.9. The distributions by treatment condition for each of these three pupil-level characteristics can be found in Appendix D: Baseline equivalence.

Table 14: Baseline characteristics of groups as analysed

School-level (continuous)	National-level mean	Intervention group		Control group		
		n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)	
School proportion English as an additional language (EAL)	16.0%	37/44 (7)	18.6 (26.4)	29/45 (16)	11.4 (16.3)	
School proportion eligible for free school meals (FSM)	25.0%	37/44 (7)	24.7 (16.4)	29/45 (16)	23.2 (16.0)	
Year 5 proportion EAL in the school	NA	37/44 (7)	16.0 (25.9)	29/45 (16)	11.5 (16.5)	
Year 5 proportion FSM in the school	NA	37/44 (7)	27.2 (20.3)	29/45 (16)	25.6 (18.3)	
KS2 reading average scaled score	104.4	36/66 (8)	104.2 (2.9)	27/66 (18)	103.7 (3.3)	
Pupil-level (categorical)		n/N (missing)	Count (%)	n/N (missing)	Count (%)	
English as an additional language status						
Non EAL pupil	84.0	691/850 (3)	81.6	532/609 (5)	88.1	
EAL pupil	16.0	156/850 (3)	18.4	72/609 (5)	11.9	
FSM status						
Non FSM pupil	75.0	628/850 (2)	74.1	457/609 (5)	75.7	
FSM pupil	25.0	220/850 (2)	25.9	147/609 (5)	24.3	
Gender						
Male	52.6	404/850 (53)	50.7	265/609 (28)	45.6	
Female	47.4	393/850 (53)	49.3	316/609 (28)	54.4	
Pupil-level (continuous)		n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)	Effect size [95% CI]
Age (in months)	NA	850/850 (0)	113.5 (3.6)	609/609 (0)	114.1 (3.6)	NA
PiRA age standardised score	NA	850/850 (0)	103.0 (14.2)	609/609 (0)	103.8 (14.0)	0.06 [-0.05 – 0.16]
Self-efficacy survey score	NA	795/850 (55)	102.0(24.0)	577/609 (32)	102.9 (21.6)	0.04 [-0.07 – 0.15]

Given that the overall attrition rate in this trial was high and that the attrition rate differed substantially across the control and PALS-UK groups (see section on Attrition for more information), it is useful to briefly explore key differences in the sample as randomised and the sample as analysed as any differences may have implications for the interpretation of the impact analysis findings.

Looking at school-level characteristics first, the tables presented in this section show that the average proportion of EAL pupils, both at the whole-school level and among Year 5 only, increased between randomisation and analysis among the sample of PALS-UK schools (from 16.6% to 18.6% and from 14.7% to 16.0% respectively) but fell substantially among the sample of control schools (from 16.2% to 11.4% and from 16.3% to 11.5% respectively). Little change was observed when comparing school-level and Year 5 proportions of FSM-eligible pupils between the samples as randomised and analysed in the PALS-UK group, while a modest decrease in these proportions was observed in the control group (from 25.1% to 23.2% and from 27.6% to 25.6% respectively). Little change was observed between randomisation and analysis when looking at the school-level KS2 average scaled score across the PALS-UK and control group.

At the pupil level, the proportion of EAL pupils in the PALS-UK group increased slightly between randomisation and analysis (from 16.8% to 18.4%) and fell substantially in the control group (from 16.6% to 11.9%). The proportion of FSM-eligible pupils in the PALS-UK group remained roughly the same between randomisation and decreased in the control group (from 27.5% to 24.3%). Similarly, the gender composition of pupils in the PALS-UK group did not change much from randomisation to analysis, while the proportion of female pupils increased between randomisation and analysis in the control group (from 51.8% to 54.4%). By age in months, very little change was observed between the samples as randomised and analysed, which was the case across the PALS-UK and control groups. The average age standardised PiRA score at baseline increased slightly from randomisation to analysis in the PALS-UK group (from 102.7 to 103.0), while a slightly bigger increase on average was observed among pupils in the control group (from 103.0 to 103.8). Finally, on the average self-efficacy survey scores, a very modest decrease in the average score among pupils in the PALS-UK group was observed between randomisation and analysis (from 130.9 to 130.6), while a modest increase was observed among the pupils in the control group between randomisation and analysis (from 131.0 to 131.7).

Taken together, these results indicate that, between randomisation and analysis, the characteristics of the PALS-UK group did not change much. The main noticeable difference in PALS-UK schools was an increase in the average proportion of EAL pupils at the school level and Year 5 level as well as an increase in the proportion of EAL pupils in the sample, between randomisation and analysis. More noticeable changes were observed in the control group, which is perhaps not surprising given the higher rate of attrition in this arm. In short, these changes included: lower school-level and Year 5 proportion of EAL pupils in the final sample; lower school-level and Year 5 proportion of FSM-eligible pupils in the final sample; lower proportion of EAL pupils in the final sample; lower proportion of FSM-eligible pupils in the final sample; higher proportion of female pupils in the final sample; higher average baseline PiRA age standardised score in the final sample; and higher average baseline self-efficacy score in the final sample. Some of these findings with regards to the control group (such as an increase in average baseline scores on the PiRA and self-efficacy survey) may indicate that the final analytical sample achieved in the control group may have performed better than what would have been observed if the attrition rate was lower and the characteristics of the final analytical sample more closely resembled the characteristics observed in the control group at randomisation.

Outcomes and analysis

Primary analysis

This evaluation uses as a primary outcome measure the age standardised score on the Progress in Reading Assessment (PiRA) test (Hodder Education). The PiRA is used as a measure of the reading skills of pupils. The PiRA Autumn 5 was used for outcome testing at baseline, while the Summer 5 version was used for outcome testing at endline. This assertion is supported by Figure 2, which indicates that a ceiling effect was not present in the endline age standardised PiRA scores.

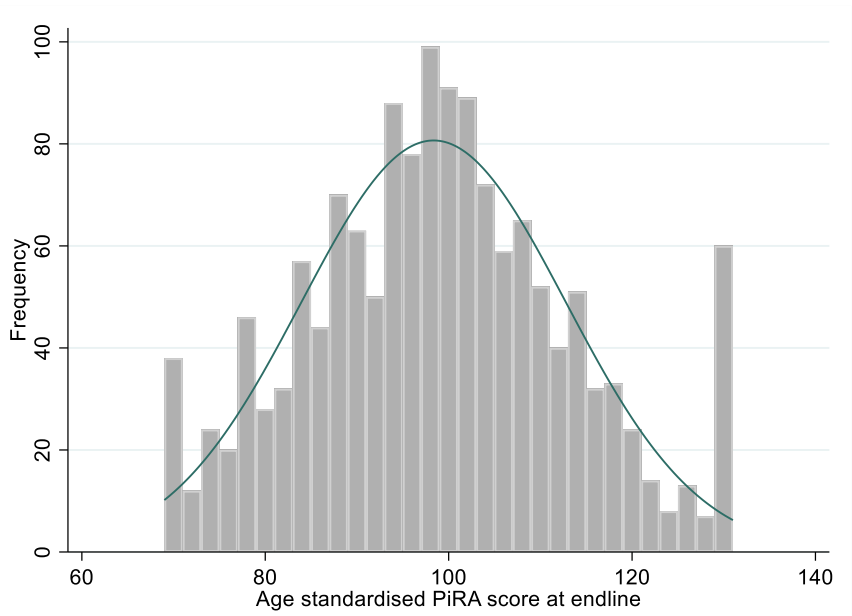
The PiRA test was administered at baseline and endline. At baseline, the PiRA was administered to 2,283 pupils across all 89 schools; at endline, 1,459 pupils across 66 schools were administered the test. For pupils included in the final analytical sample (1,459 with observable data for the age standardised score on the PiRA test taken at baseline and endline), the age standardised score at endline had an overall mean of 98.4 and a standard deviation of 14.4.

The PiRA marking guidelines, as set out in the PiRA manual, require that all scores below 70 are bottom coded as '<70', while scores above 130 are top coded as '>130'. We recoded these top and bottom-end scores such that scores of <70 became 69 and scores of >130 became 131 in order to facilitate the conversion of the data to numeric to facilitate the analysis The range of the analytical sample was therefore from 69 (the score assigned to all pupils with an age standardised score below 70) to 131 (the score assigned to all pupils with an age standardised score above 130). Thus, no detail in the scores was lost by recoding the data in this way, although this approach to scoring is more likely to produce ceiling and floor effects, which can be observed to some extent in Figure 2 and Figure 3. Among the baseline age standardised PiRA scores as presented in Figure 3, 2.3% of pupils (34 out of 1,459) had a score of <70, while 1.2% of pupils (18 out of 1,459) had a score of >130. Top and bottom-end scores by treatment group did not vary much from the overall figures at baseline, with 2.0% of pupils (12 out of 609) in the control group scoring <70 and 1.6% (10 out of 609) scoring >130, while 2.6% of pupils (22 out of 850) scored <70 and 0.9% of pupils (8 out of 850) scored >130 in the PALS-UK group. At endline, 2.1% of pupils (31 out of 1,459) achieved a score of <70, while 3.6% of pupils (52 out of 1,459) achieved a score of >130. By treatment group, at endline the proportion of top and bottom-end scores did not vary much from the overall figures either, with 2.3% (14 out of 609) of pupils scored <70 and 3.9% of pupils (24 out of 609) scored >130 in the control group, while 2.0% of pupils (17 out of 850) scored <70 and 3.3% of pupils (28 out of 850) scored >130 in the PALS-UK group.

The distribution of the primary outcome measure is illustrated in Figure 2. The distribution of age standardised PiRA scores at baseline and endline by treatment group can be found in Appendix E. These figures show that the distribution of baseline and endline PiRA scores were roughly similar across the PALS-UK and control groups.

The spread of the outcome measure closely resembles a normal distribution indicating that the assumptions underlying the parametric tests used in the analysis are not violated. Nonetheless, the parametric tests used in the analysis of data are robust to any slight skew in the data and therefore the variable as shown below was used without any further modification or transformation as the outcome variable for the primary analysis. A higher score on this variable indicates a better reading attainment outcome.

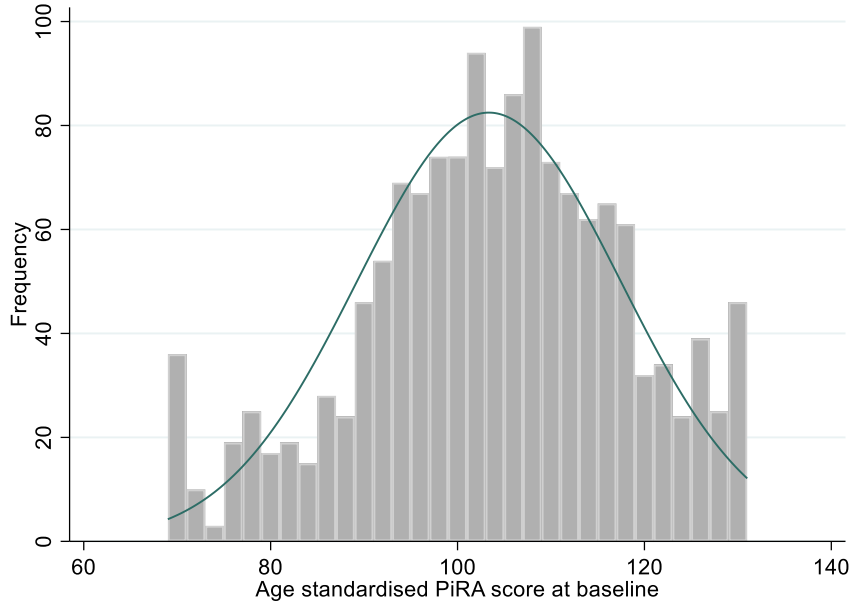
Figure 2: Histogram showing the distribution of age standardised PiRA scores at endline



This histogram includes only the 1,459 pupils that were included in the final analytical model.

The primary outcome measure was correlated to the language baseline measure (age standardised score on the PiRA), with r = 0.63. The distribution of age standardised PiRA scores at baseline is shown in Figure 3.

Figure 3: Histogram showing the distribution of age standardised PiRA scores at baseline



This histogram includes only the 1,459 pupils that were included in the final analytical model.

As outlined in the

Methods section, the primary analysis took an intention to treat (ITT) approach, with both stratification variables included in the randomisation also present in the analysis (that is, geographical region and whether the school was a single or multi-form entry). Multilevel models were used to assess the effectiveness of the PALS-UK intervention so as to account for the clustering of children in schools. The details of the model for the primary outcome analysis are reported in Table 15 and Table 16.

The unadjusted mean for the intervention group was lower than the unadjusted mean for the business-as-usual control group at endline. As shown in Table 15, the unadjusted difference in means was -1.239. In the multilevel model that accounted for the clustering of pupils in schools and also accounted for PiRA baseline scores, geographical region and single- or multi-form entries an adjusted difference in means of -0.834 age standardised scores on the PIRA was observed (Table 16). The lower bound 95% interval estimate here was -3.136 age standardised scores, while the upper bound 95% interval estimate was 1.468. This equates to an effect size point estimate of -0.07, which in turn suggests that pupils in the PALS-UK group had made one month less progress, on average, as compared to those in the control group. The lower bound 95% interval for the effect size estimate was -0.29, which equates to four months less progress among pupils in the PALS-UK group as compared to pupils in the control group, while the upper bound 95% interval estimate was 0.14, which equates to two months' additional progress made by pupils that received the PALS-UK programme. The p-value associated with the difference in means was p = 0.478, suggesting that there is a relatively high probability that the model estimates were observed due to statistical uncertainty. Thus, the results here suggest that reading attainment as measured by the PiRA was similar across the PALS-UK and control group at endline.

Table 15: Primary outcome analysis results

			Intervention group		Control group		
Outcome	Unadjusted differences in means (I-C)	Adjusted differences in means (95% CI)	n (missing)	Variance of outcome	n (missing)	Variance of outcome	Pooled variance
Primary outcome: age standardised PiRA score	-1.239	-0.834 (-3.136–1.468)	850 (263)	192.740	609 (460)	229.558	208.246

Table 16: Effect size estimation, primary outcome analysis

	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
Primary outcome: age standardised PiRA score	850 (263)	97.827 (96.2–99.4)	609 (460)	99.066 (96.4–101.7)	1,459 (850; 609)	-0.07 (-0.29–0.14)	0.478

In conclusion, the primary analysis did not identify any conclusive positive or negative impact of PALS-UK on the reading attainment of Year 5 pupils in the sample due to the high level of statistical uncertainty in the results. While the best estimate suggests that pupils who received PALS-UK had made, on average, one month less progress in reading attainment, the relatively high p-value suggests that there was high probability that the model estimates were observed due to statistical chance or uncertainty.

Secondary analyses

The updated protocol and statistical analysis plan set out one secondary outcome measure: self-efficacy in reading. In addition, we report descriptive findings based on a small sample of schools that administered the reading comprehension subtest. The results of these analyses are presented here in turn for the two outcomes.

Self-efficacy in reading

The secondary outcome measure in this trial was pupil's self-efficacy in reading, as assessed through the self-efficacy survey administered at baseline and endline. Endline data was available for 1,372 pupils from 62 schools, while the baseline survey was collected from 2,166 pupils from 82 schools**.** The extent of the missing data (equal to 36.7% attrition) introduces bias into the design and reduces the extent to which we can make precise estimates, particularly given the relatively small sample size of the populations.

For the overall analytical sample, self-efficacy in reading scores had a mean of 102.37 and standard deviation of 23.02 at baseline and a mean of 101.19 and standard deviation of 21.62 at endline. The distribution on the survey is provided in Figure 4 and Figure 5, slightly skewed to the upper end of the range of scores at both time points.

Figure 4: Secondary outcome measure—self-efficacy in reading at baseline

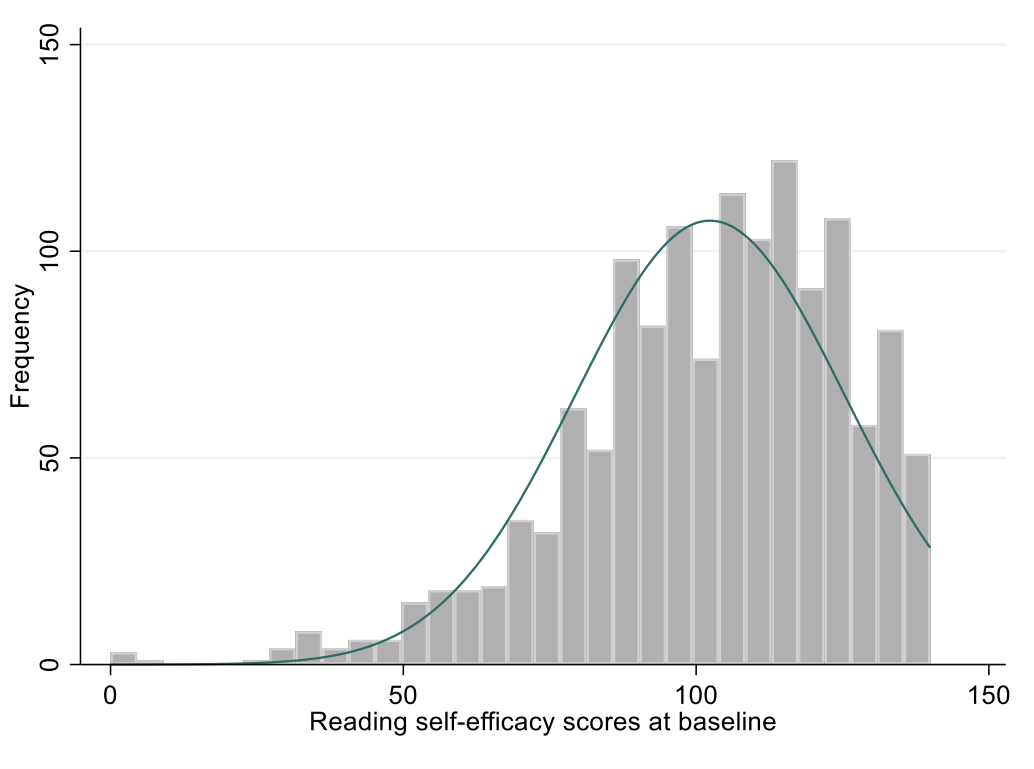
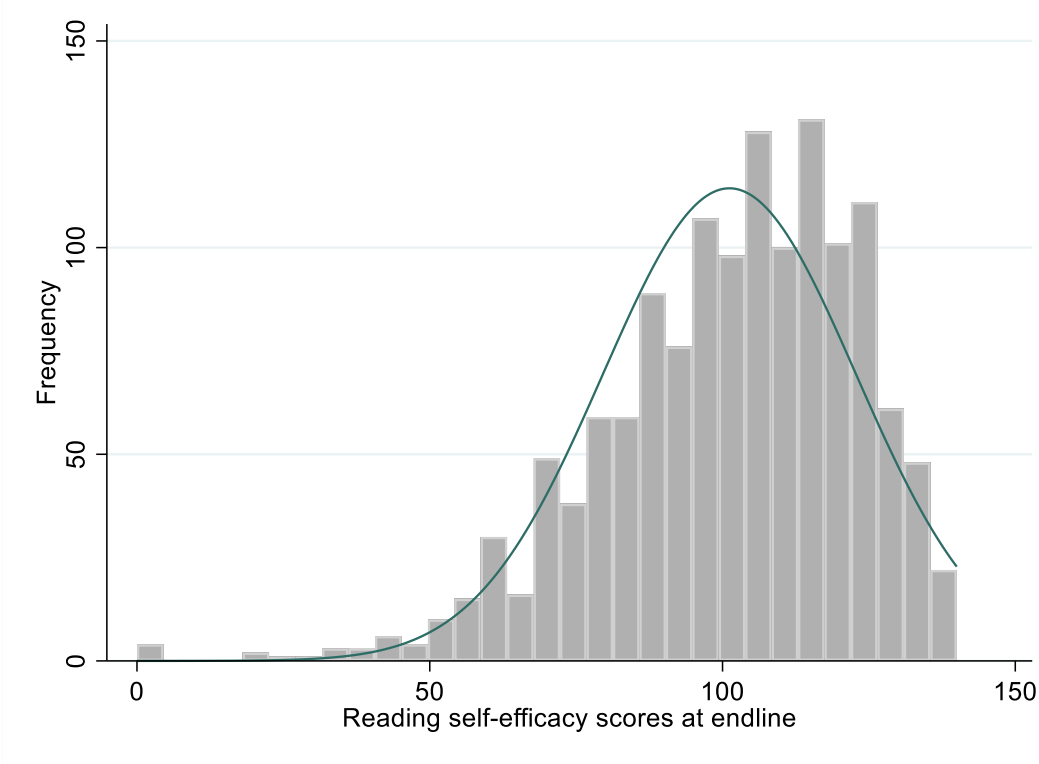


Figure 5: Secondary outcome measure—self-efficacy in reading at endline



We compared treatment and control classes using the same model specification as for the primary outcome, substituting the secondary outcome and using baseline self-efficacy scores instead of reading attainment. A higher score on the self-efficacy survey indicates higher confidence in reading.

For self-efficacy, there is almost no difference in the unadjusted means between the PALS-UK group (mean = 100.986; 95% CI: 98.050; 103.922) and the control group (mean = 101.529; 95% CI: 98.709; 104.349) at endline (Table 17). After controlling for pre-intervention covariates, the adjusted difference in mean between the intervention and control group was -0.068, while the effect size associated with this difference was 0.00 or equivalent to zero months' additional progress in reading self-efficacy in the PALS-UK group (95% CI: -0.20; 0.20), with a significance level of 0.973 (Table 18). This indicates that pupils' reading self-efficacy was similar in both treatment and business-as-usual schools at endline.

Table 17: Secondary outcome analysis—self efficacy in reading

			Intervention group		Control group		
Outcome	Unadjusted differences in means (I-C)	Adjusted differences in means (95% CI)	n (missing)	Variance of outcome	n (missing)	Variance of outcome	Pooled variance
Secondary outcome: self-efficacy in reading	-0.583 (-4.751; 3.586)	-0.068 (-4.000; 3.863)	795 (297)	484.83	577 (441)	443.81	467.26

Table 18: Effect size estimation, secondary outcome analysis—self efficacy in reading

	Intervention group	Control group	Effect size

Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention; control)	Hedges g (95% CI)	p-value
Secondary outcome: self-efficacy in reading	795 (297)	100.986 (98.050; 103.922)	577 (441)	101.529 (98.709; 104.349)	1,372 (795; 577)	0.00 (-0.20; 0.20)	0.973

Overall, then, the secondary analysis yielded no conclusive evidence that PALS-UK has any positive or negative impact on reading self-efficacy.

Reading comprehension—WIAT-III-UK-TI

As discussed in the updated evaluation protocol and SAP, the WIAT-III-UK-T element of the outcome testing was only partially administered due to the COVID-19 pandemic: data for only 18 schools (11 control; 7 PALS-UK) was collected for the WIAT-III-UK-T reading comprehension subtest and none of the schools completed the WIAT-III-UK-T reading fluency subtest. Therefore, the analysis is not able to detect relevant differences. For this reason, we do not report on the results in the main body of the analysis to make it clear that any analysis is presented only for descriptive and informational purposes. Instead, descriptive analyses of the sample of pupils that were administered the WIAT-III-UK-T reading comprehension subtest are presented in Appendix I.

Analysis in the presence of non-compliance

In total, 21 of the 37 schools for which endline outcome data was available were compliant, which suggests that the results from the compliance analysis may differ from the ITT analysis. It should be noted that non-compliance here is predominantly driven by lower delivery survey response rates among schools in the PALS-UK group, upon which compliance criteria 2 and 3 were based (Table 6).

The results of the compliance analysis using the binary compliance measure outlined in the

Methods section (see **Analysis in the presence of non-compliance**) suggest that there was no conclusive evidence of any positive or negative impact of compliance with the PALS-UK intervention due to the high level of statistical uncertainty in the results. While a small, non statistically significant impact of increased compliance on pupils reading outcomes was observed (coefficient for binary fidelity variable after 2SLS = 0.893; p = 0.577; N = 1,459, 95% CI: -2.244; 4.031 (Table 19) and the direction of the effect was positive, it was very small in magnitude—0.04, 95% CI: -0.22; 0.29) suggesting that there are no differential treatment effects among schools with high levels of compliance (equivalent to zero months' additional progress in the PALS-UK group). We note that limitations in our approach here are the low survey response rate and, linked to this, that compliance is not precisely addressed with the survey questions as the non-responses and missing data from the surveys were coded as zero i.e. non-compliant.

Table 19: Analysis in the presence of non-compliance—main outcome

	Compliance measure 1 (main)					
Outcome	Coefficient (CI)	Standard error	Z	p value	N	Effect size (95% CI)
PIRA: high compliers group	0.893 (-2.244; 4.031)	1.601	0.56	0.577	1,459	0.04 (-0.22; 0.29)

In addition, we ran the same model using alternative compliance specifications. The first stage results of the additional compliance analysis using compliance measure 2 and compliance measure 3 are presented in Table 20 and Table 21.

The results of the compliance analysis using these different specifications of the compliance measure suggest that there was no conclusive evidence of any positive or negative impact of increased compliance with the PALS-UK intervention due to the high level of statistical uncertainty in the results.

Table 20: Analysis in the presence of non-compliance—additional compliance measure 2

	Compliance measure 2					
Outcome	Coefficient (CI)	Standard error	Z	p value	N	Effect size (95% CI)
PiRA	1.514 (-1.533, 4.56)	1.554	0.97	0.330	1,459	0.05 (-0.19; 0.28)

Table 21: Analysis in the presence of non-compliance—additional compliance measure 3

	Compliance measure 3					
Outcome	Coefficient (CI)	Standard error	Z	p value	N	Effect size (95% CI)
PiRA	1.864 (-1.49; 5.221)	1.713	1.09	0.276	1,459	0.03 (-0.19; 0.26)

Missing data analysis

Before proceeding with the missing data analysis, it is important to clarify the effective sample size. As specified in the section on Participant flow including losses and exclusions, a total of 2,283 pupils were randomised in this PALS-UK trial. However, pupils were only considered for the missing data analysis if they had complete PiRA data at baseline, which was not necessarily the case for all randomised pupils. Given that 108 pupils were randomised but did not have baseline PiRA data, the effective sample for the missing data analysis was 2,175 pupils. Baseline data was missing for pupils due to a combination of factors such as absence on the testing day, a school error resulting in missing data, or a school not administering the baseline assessment due to pupil SEND.

Moreover, while the models that we present in this section do not distinguish between missingness explained by pupil- and school-level factors, we illustrate in the Attrition section that the main mechanism of attrition in this trial was at the school level.

Table 22: Missing PiRA data at endline

Missingness	Frequency	Percent (%)
Complete baseline and endline PiRA data	1,459	67.1
Missing endline PiRA data	716	32.9
Total	2,175	100

Note: Pupils were deemed to have missing data if they had complete baseline PiRA data but no endline PiRA data.

The overall share of missing data, as shown in Table 22, was 32.9%. Table 23 shows the distribution of missing primary outcome data by treatment allocation. Overall, 716 pupils were deemed as having missing data, with an uneven distribution across intervention and control (23.6% and 42.7% respectively).

Table 23: Missing primary outcome data across intervention and control schools

Treatment	Complete baseline and endline PiRA data	Missing endline PiRA data	Total
Control (BAU)	609 (57.3%)	453 (42.7%)	1,062
PALS-UK	850 (76.4%)	263 (23.6%)	1,113
Total	1,459	716	2,175

Pupils were deemed to have missing data if they had complete baseline PiRA data but no endline PiRA data.

Analysis of Table 24 shows that missingness at a descriptive level appeared to be associated with a slightly lower mean age standardised score on the PiRA at baseline. This was the case when analysing the mean age standardised PiRA scores at baseline across the overall sample by missing data status, and when exploring this separately across the intervention and control groups. Given that missingness in the endline data is predominantly explained at the school level (see section on Attrition for more information), this indicates that schools with slightly more lower-achieving pupils at baseline were more likely to drop out from the trial.

Table 24: Missingness by mean age standardised score on the PiRA at baseline

Missingness	Mean age standardised PiRA score at baseline (N=)		
	Intervention	Control	Overall
Complete baseline and endline PiRA data	103.0 (850)	103.8 (609)	103.4 (1,459)
Missing endline PiRA data	101.9 (263)	102.0 (453)	101.9 (716)
Overall	102.7 (1,113)	103.0 (1,062)	102.9 (2,175)

Looking at the difference now by region and number of entry type (single or multi-form), Table 25 shows that the proportion of missingness was higher among multi-form entries in the Midlands (42.1%) when compared to the other categories.

Table 25: Missing primary outcome data by geographical region and number of classes

	Complete baseline and endline PiRA data	Missing endline PiRA data	Total
Midlands 1 class	471 (78.1%)	132 (21.9%)	603
Midlands >1 class	581 (57.9%)	422 (42.1%)	1,003
North East 1 class	298 (71.0%)	122 (29.0%)	420
North East >1 class	109 (73.2%)	40 (26.8%)	149
Total	1,459	716	2,175

By FSM and EAL status at the pupil level (Table 26), analysis showed that missingness at a descriptive level was higher for these subgroups as compared to pupils not in these subgroups. Linked to this, missingness tended to be slightly higher in schools where the share of FSM and EAL pupils in Year 5 was higher also (Table 27).

Table 26: Missing primary outcome data by pupil FSM and EAL status

Missingness	Non-FSM pupils	FSM pupils	Total	Non-EAL pupils	EAL pupils	Total
Complete baseline and endline PiRA data	1,091 (68.6%)	361 (63.0%)	1,452	1,223 (67.9%)	228 (63.2%)	1,451
Missing endline PiRA data	500 (31.4%)	212 (37.0%)	712	579 (32.1%)	133 (36.9%)	712
Total	1,591	573	2,164	1,802	361	2,163

Table 27: Mean Year 5 percentage of FSM and EAL pupils by missing primary outcome data status

Missingness	Mean Year 5 % FSM pupils (N=)	Mean Year 5 % EAL pupils (N=)
Complete baseline and endline PiRA data	25.5 (1,455)	15.9 (1,455)
Missing endline PiRA data	28.3 (716)	18.2 (716)
Total	2,171	2,171

Results from the multilevel binary logistic regression model show that the odds of missing data were substantially lower among pupils in PALS-UK schools (Odds Ratio [OR] = 0.148, p = 0.005) as compared to those in control schools (all model estimates and specifications can be found in the **Error! Reference source not found.** section under Appendix G). The lower bound estimate here was 0.039, while the upper bound estimate was 0.563.

The odds of missingness were slightly lower for pupils with higher prior attainment (OR = 0.988, p = 0.059), or in other words, a one-unit increase in the age standardised score on the PiRA at baseline was associated with slightly lower odds of missingness at endline. The narrow intervals suggested that this point estimate was well supported by the data (lower bound 95% interval estimate = 0.975, upper bound 95% interval estimate = 1.000).

As indicated by the descriptive analysis, the odds of missingness were considerably higher among pupils in multi-form entry schools from the Midlands when compared to single-form entries from the Midlands (OR = 0.094, p = 0.005; lower bound 95% interval estimate = 0.018, upper bound 95% interval estimate = 0.493). The odds of missingness among pupils in multi-entry schools in the Midlands were also higher than for those pupils from single-form entries from the North East (OR = 0.162, p = 0.063; lower bound 95% interval estimate = 0.240, upper bound 95% interval estimate = 1.100). Finally, when comparing multi-form entries from the Midlands to multi-form entries in the North East, the odds of missingness among pupils from the former were again seemingly higher than among pupils from the latter (OR = 0.101, p = 0.105; lower bound 95% interval estimate = 0.006, upper bound 95% interval estimate = 1.611).

Looking at missingness by FSM status, the odds of missing data were higher among the FSM subgroup as compared to those not in the subgroup (OR = 1.458, p = 0.057; lower bound 95% interval estimate = 0.989, upper bound 95% interval estimate = 2.150). The best estimate therefore suggests that FSM pupils were 1.5 times more likely to have missing primary outcome data when compared to pupils that were not receiving FSM, with the upper bound estimate indicating two times more likely. When interacted with treatment allocation, the odds of missing outcome data were substantially lower among FSM pupils in PALS-UK schools as compared to FSM pupils in control schools (OR = 0.137, p = 0.007; lower bound 95% interval estimate = 0.032, upper bound 95% interval estimate = 0.585).

There was some evidence to suggest that EAL status was associated with higher odds of missingness in the primary outcome (OR = 1.474, p = 0.191; lower bound 95% interval = 0.824, upper bound 95% interval = 2.639). This suggests that pupils speaking EAL were 1.5 times more likely to have missing primary outcome data as compared to pupils speaking English as their first language. Furthermore, when EAL status was interacted with treatment allocation, stronger evidence of a pattern related to EAL status in the missing primary outcome data was observed. Indeed, EAL pupils in PALS-UK schools had significantly lower odds of missing primary outcome data as compared to EAL pupils in control schools (OR = 0.148, p = 0.022; lower bound 95% interval estimate = 0.029, upper bound 95% interval estimate = 0.756).

Finally, looking at Year 5 percentage of FSM and EAL pupils in the school, neither appeared to have any bearing on missing primary outcome data at the pupil level. First, looking at Year 5 percentage of FSM pupils, the odds of missingness did not change with increasing proportion of Year 5 FSM pupils in the school (OR = 1.003, p = 0.872; lower bound 95% interval estimate = 0.967, upper bound 95% interval estimate = 1.041). Similar results were observed when looking at percentage of Year 5 EAL pupils in the school, with a in the likelihood of missingness in the primary outcome data observed with an increase in the percentage of Year 5 EAL pupils (OR = 1.003, p = 0.842; lower bound 95% interval estimate = 0.974, upper bound 95% interval estimate = 1.033).

Given that the proportion of missing primary outcome data exceeded 5%, we pursued an analytical approach to understand if the data appeared to be missing completely at random (MCAR) or whether the weaker Missing at Random (MAR) assumption applies as specified in the statistical analysis plan. To do this, we first implemented Little's test of MCAR to assess if the data was consistent with being MCAR. The results of this analysis are presented in Appendix , under missing data analysis. The null hypothesis here was that the data was MCAR. However, as shown in the model outputs in Appendix , we can reject this null hypothesis, with the probability that these results were observed due to statistical uncertainty being very low (P < 0.0001).

As specified in the statistical analysis plan, we then ran a pattern mixture model. The aim was to model data and missingness jointly as the main treatment effect was estimated. Key in running a pattern mixture model is to input assumptions about the mean difference in outcomes between pupils that were unobserved at endline (U) and those that were (O), i.e. U-O. In the context of pattern mixture modelling this is referred to as specifying the delta. In the inherent absence of data to make these assumptions on the basis of endline age standardised PiRA scores, we compared the mean baseline age standardised PiRA scores of those that were unobserved at endline with those that were observed, i.e. using the means as presented in Table 24. Those that were observed at endline had a higher mean age standardised score at baseline (103.4) than the unobserved pupils (101.9), representing a difference of -1.5 percentage points. Based on this, we ran the pattern mixture model, which was based on the primary analysis model, with delta specified as -1.5. The results of the pattern mixture model are presented in Table 28.

Table 28: Pattern mixture model results

Outcome	Adjusted differences in means (95% CI)	Intervention group	Control group	p-value
		n (missing)	n (missing)	
Primary outcome: age standardised PiRA score	-0.193 (-3.031–2.645)	1,113 (60)	1,062 (48)	0.892

N = 2,175, comprised of 1,459 pupils (PALS-UK: 850; control: 609) with observed outcomes and 716 pupils (PALS-UK: 263; control: 453) with unobserved outcomes. The 108 missing pupils in this model are those from the entire sample of 2,283 pupils that did not have an age standardised PiRA score at baseline, as explained in the Participant flow including losses and exclusions section.

The results of this missing data analysis are consistent with the findings of the primary analysis, indicating no conclusive evidence of any positive or negative impact of PALS-UK on reading attainment when observed and missing data are modelled together. An adjusted difference in means of -0.193 was observed (Table 28). The lower bound 95% interval estimate here was -3.031, while the upper bound estimate was 2.645, indicating that data supported a wide range of scenarios ranging from negative impact of PALS-UK as well as positive impact on the primary outcome. The best estimate of the impact of PALS-UK in this trial using the pattern mixture model suggests that there was almost no difference in outcomes between unobserved and observed pupils in the control group as compared to their counterparts in the PALS-UK group. However, the large p-value here (p = 0.892) indicates that that there was a high probability that the model estimates were observed due to statistical uncertainty.

Subgroup analyses

We pre-specified in the SAP that we would assess subgroup effects for three separate groups: (1) FSM-eligible pupils, (2) pupils with EAL, and (3) pupils with low reading attainment. The subgroup analysis was undertaken for the primary outcome measure only. As mentioned in the SAP, all subgroup analyses in this evaluation were exploratory and likely underpowered.

Subgroup analysis—FSM group

The hypothesis tested was that PALS-UK had differential effects for FSM pupils in the treatment group compared to FSM pupils in the control condition. At endline, there were 24.86% pupils (366 out of 1,452) that were identified as FSM: 25.94% (220 out of 847) of analytical sample pupils in intervention schools and 24.17% (146 out of 604) in control schools. For the analytical sample of FSM pupils, reading attainment measured with the PiRA had a mean of 93.99 and standard deviation of 13.69 at endline for schools in the treatment groups and a mean of 91.721 and standard deviation of 13.73 for schools in the control group.

The unadjusted mean (2.269) was over two age standardised scores higher among pupils that received PALS-UK as compared to those in the control group. The estimate of the interaction coefficient in the model suggests an overall effect of PALS-UK (an adjusted difference in means) for FSM pupils in intervention compared to control schools of 0.738 of an age standardised score, as indicated in Table 29. The lower bound 95% interval estimate here was -2.250, while the upper bound estimate was 3.726. The point estimate equates to an effect size of 0.054 or one month's additional progress in reading (see Table 30). However, the data was also compatible with a range of other scenarios, with the lower bound effect size estimate at -0.16, while the upper bound estimate was 0.27. The p-value here (p = 0.63) suggests that there was a high probability that the results were observed due to statistical uncertainty.

We also undertook the FSM subgroup analysis using the FSM-eligible sample only (that is, as a separate sub-sample). The results are presented in Appendix G. The overall pattern of the results holds for the FSM sample, leading to the conclusion that there are no detectable differences in the impact of PALS-UK on the reading attainment of pupils in the FSM subgroup.

Notably, the previously negative coefficient on PiRA observed in the primary outcome analysis is now positive, however, despite the positive difference, we cannot reject the hypothesis that PALS-UK had no effect on the FSM subgroup. While the FSM intervention group obtained slightly higher PiRA results compared to the control group, the data also supported scenarios where FSM pupils allocated to the control condition obtained slightly higher PiRA scores, and there is evidence that the best estimates were likely observed due to statistical uncertainty. The computation of the effect sizes confirms this assessment; this additionally controls for prior PiRA attainment. Note that for some Year 5 pupils there is no valid FSM indicator or they could not be merged (n = 7) and so they could not be included in the model.

Overall, this exploratory and underpowered analysis did not find any conclusive evidence of differential impacts of PALS-UK on the reading attainment of FSM pupils. Indeed, while FSM pupils in the PALS-UK group made one additional month's progress, the large p-values suggests that there is a high probability that these results were observed due to statistical chance or uncertainty.

Table 29: Subgroup analysis—FSM pupils

	Intervention group: FSM Subgroup	Control group: FSM Subgroup	

Outcome	Unadjusted differences in means	Adjusted differences in means (95% CI)	n (missing)	Variance of outcome	n (missing)	Variance of outcome	Pooled variance
Primary outcome: FSM group (interaction)	2.269	0.738	220 (67)	187.47	147 (147)	188.50	188.611

Table 30: Effect size calculation: subgroup analysis—FSM pupils

	Unadjusted means				Effect size		
	Intervention group: FSM subgroup		Control group: FSM subgroup				
Outcome	N (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention; control)	Hedges g (95% CI)	p-value
Primary outcome: FSM group (interaction)	220 (6867)	93.991 (92.171; 95.810)	147 (150147)	91.721 (89.483; 93.959)	1,451 (847, 604)	0.05 (-0.16; 0.27)	0.63

Subgroup analysis—EAL group

We also examined EAL subgroup effects. At endline, 15.71% of pupils (228 of 1,452) were identified as EAL:18.42% (156 of 847) of analytical sample pupils in intervention schools and 11.92% (72 of 604) of analytical sample pupils in control schools. The results reflected the ITT analysis above: there were no differences between control and treatment EAL pupils. The results are presented in Table 31 and Table 32.

The results estimated by interacting EAL status with treatment allocation point to no conclusive effect of PALS-UK on the reading attainment as measured by the PiRA test. Indeed, while the unadjusted difference in means was over two age standardised scores on the PiRA (2.075), the adjusted difference in means between EAL pupils that received PALS-UK and those allocated to the control condition was 0.626 of an age standardised score on the PiRA. The lower bound 95% interval estimate here was -3.247, while the upper bound estimate was 4.498, indicating that the data supported a wide range of scenarios with regards to the impact of PALS-UK on the reading attainment of EAL pupils. The point estimate of the average treatment effect equivalises to an effect size of 0.043, or zero additional months' progress in reading (see Table 32). Again, the data supported a range of scenarios with regards to the effect of PALS-UK on the reading attainment of EAL pupils; the lower bound effect size estimate was -0.22, while the upper bound effect size estimate was 0.31. A high p-value was observed when estimating this model (P = 0.75) indicating a strong possibility that the estimates observed here were due to statistical uncertainty. Taking these results together, we cannot reject the hypothesis that PALS-UK had no effect on EAL pupils.

Table 31: Subgroup analysis—EAL pupils

	Intervention group: EAL Subgroup	Control group: EAL Subgroup	

Outcome	Unadjusted differences in means	Adjusted differences in means (95% CI)	n (missing)	Variance of outcome	n (missing)	Variance of outcome	Pooled variance
Primary outcome: EAL group (interaction)	2.075	0.626 (-3.247–4.498)	156 (30)	219.32	72 (103)	210.39	216.4964

Table 32: Effect size calculation: subgroup analysis—EAL group

	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
Primary outcome: EAL group (interaction)	156 (3130)	98.244 (95.901; 100.586)	72 (105103)	100.319 (96.911; 103.728)	1,451 (847, 604)	0.04 (-0.22; 0.31)	0.75

The results for the model estimated by including only the EAL sample are presented in Appendix G. This alternative analysis confirms that there is no conclusive evidence to suggest that PALS-UK had any positive or negative impact on the reading attainment of EAL pupils.

Overall, then, the results of the EAL subgroup analysis indicate that PALS-UK did not have any conclusive positive or negative impact on the reading attainment of this subgroup. This finding was robust to two specifications of the model: interacting EAL status with treatment allocation, and running the primary outcome analysis model on the isolated subsample of EAL pupils.

Subgroup analysis—pupils with low and high reading attainment

As set out in the trial protocol and SAP, we ran subgroup analyses for the following subgroups:

- pupils with lower reading attainment at baseline—pupils with below average age standardised scores on the PiRA test at baseline;
- pupils with higher reading attainment at baseline—pupils with average or above age standardised scores on the PiRA test at baseline;
- pupils with very low reading attainment at baseline—pupils in the bottom quartile for age standardised scores on the PiRA test at baseline; and
- pupils with very high reading attainment at baseline—pupils in the top quartile for age standardised scores on the PiRA test at baseline.

As specified in the

Methods section, binary variables to identify below/above average readers at baseline were constructed based on the sample-specific mean PiRA score to ensure as much as possible that even samples could be achieved for the analysis.

This aspect of the subgroup analysis informed if there are any differences in programme effect for low and very low ability readers as compared to higher ability pupils. Specifically, the analyses tested the hypothesis that PALS-UK had a positive effect on the outcomes of low and very low ability readers in intervention schools as compared to those in business-as-usual control schools. FSM pupils are overrepresented in the pupils with very low reading attainment category (35.4% of FSM pupils were very low attainers in reading at baseline) and low reading attainment (30.8%) and underrepresented in the pupil with very high reading attainment category (12.0%) (see Table 33). There was also representation of FSM pupils in the high reading attainment category (21.8%). Thus, while FSM pupils were more likely to have lower reading attainment at baseline, they were also represented in the higher attaining groups at baseline. As a result, we would not necessarily expect to observe the same findings around the impact of PALS-UK for FSM pupils and pupils with lower reading attainment at baseline.

Table 33: Share of FSM pupil that are classified as pupils with very low, low, high, and very high reading ability based on the PiRA score at baseline

	Quartile 1: Very low attainers			Quartile 2: Low attainers			Quartile 3: High attainer			Quartile 4: Very high attainers		
FSM eligible	T	C	Total	T	C	Total	T	C	Total	T	C	Total
Count	76	54	130	68	45	113	49	31	80	27	17	44
Share	34.55	36.73	35.42	30.91	30.61	30.79	22.27	21.09	21.80	12.27	11.56	11.99

Overall, due to the high level of statistical uncertainty around these results, this part of the impact evaluation also provides no conclusive evidence of any positive or negative impact of PALS-UK among pupil subgroups with varying levels of reading attainment at baseline. Looking first at pupils with lower reading attainment at baseline, Table 34 reports the results from the multilevel model and effect size estimation. At endline, there were 542 pupils with below average reading attainment as measured by the PiRA at baseline: 37.88% of pupils in the primary analytical sample fell into this category in intervention schools; this was 36.12% in control schools.

Table 34: Subgroup analysis results—pupils with lower reading attainment at baseline

			Intervention group		Control group		
Outcome	Unadjusted differences in means (I-C)	Adjusted differences in means (95% CI)	n (missing)	Variance of outcome	n (missing)	Variance of outcome	Pooled variance
Age standardised PiRA score: lower reading attainment	-1.460	-1.153 (-3.892 - 1.586)	322 (0)	137.461	220 (0)	165.449	148.550

Table 35: Effect size estimation, subgroup analysis—pupils with lower reading attainment at baseline

	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
Age standardised PiRA score: lower reading attainment	322 (0)	88.909 (86.965–90.854)	220 (0)	90.369 (87.428–93.310)	542 (322; 220)	-0.09 (-0.32–0.13)	0.409

The unadjusted mean for the intervention group was once again lower for the control group. As shown in Table 34, the unadjusted difference in means was -1.460. In the multilevel model that accounted for the clustering of pupils in schools and also accounted for PiRA baseline scores, geographical region, and single or multi-form entry, the adjusted difference in means was -1.153 age standardised scores on the PiRA. However, the data supported a range of scenarios here (lower bound 95% interval estimate -3.892, upper bound estimate 1.586). The point estimate here equivalises to an effect size of -0.09, or one month less progress for lower attaining pupils that received PALS-UK (Table 35). Again, the data supported a range of scenarios ranging from a lower bound effect size estimate of -0.32 to an upper bound estimate of 0.13. This variability in the effect size compatibility interval indicates a higher level of uncertainty in these results, which is affirmed by the relatively high p-value of 0.409.

Focusing now on those with higher reading attainment at baseline, Table 36 and Table 37 report the results from the multilevel model and effect size estimation. At endline, there were 917 pupils with average or above reading attainment as measured by the PiRA at baseline: 62.12% of pupils in the primary analytical sample fell into this category in intervention schools while this was 63.88% of pupils in control schools.

Table 36: Subgroup analysis results—pupils with higher reading attainment at baseline

			Intervention group		Control group		
Outcome	Unadjusted differences in means (I-C)	Adjusted differences in means (95% CI)	n (missing)	Variance of outcome	n (missing)	Variance of outcome	Pooled variance
Age standardised PiRA score: higher reading attainment	-0.753	-0.647 (-3.119–1.826)	528 (0)	150.860	389 (0)	175.766	161.750

Table 37: Effect size estimation, subgroup analysis—pupils with higher reading attainment at baseline

	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
Age standardised PiRA score: high reading attainment	528 (0)	103.565 (101.707–105.424)	389 (0)	104.318 (102.180–106.457)	917 (528; 389)	-0.05 (-0.25–0.14)	0.608

The unadjusted mean for the intervention group was once again lower than that of control group. As shown in Table 36, the unadjusted difference in means was -0.753.

In the multilevel model that accounted for the clustering of pupils in schools and also accounted for PiRA baseline scores, geographical region, and single or multi-form entry, a similar but not statistically significant difference in means was observed with p = 0.608. The adjusted difference in means obtained from the multilevel model was -0.647. The effect size associated with this adjusted difference in means is -0.05, equivalent to one months' less progress for pupils in the PALS-UK group (Table 39).

This part of the subgroup analysis suggests that there was no conclusive impact of PALS-UK on the reading attainment of pupils with higher attainment in reading at baseline. The data supported a wide range of scenarios, with the lower bound 95% interval estimate being -3.119 age standardised scores, while the upper bound estimate was 1.826. The point estimate equivalises to an effect size of -0.05, or one month less progress in reading among those higher attainers allocated to the PALS-UK group. Again, the data supported a wide range of scenarios here, with the lower bound effect size estimate at -0.25, while the upper bound estimate was 0.14. The high p-value here (P=0.608), however, indicates that there was a high probability that the estimates observed were due to statistical uncertainty.

Moving on to focus on pupils with very low attainment at baseline, Table 38 and Table 39 present the results of the multilevel model and effect size estimation. At endline, there were 365 pupils in the overall sample scoring in the bottom quartile for reading attainment at baseline as measured by the PiRA: 25.18% of the primary analytical sample were categorised as having very low reading attainment at baseline in intervention schools; in control schools it was 24.79%.

Table 38: Subgroup analysis results–very low ability readers at baseline

			Intervention group		Control group		
Outcome	Unadjusted differences in means (I-C)	Adjusted differences in means (95% CI)	n (missing)	Variance of outcome	n (missing)	Variance of outcome	Pooled variance
Age standardised PiRA score: very low ability readers	-2.236	-2.025 (-5.048–0.997)	214 (0)	121.262	151 (0)	170.914	141.506

Table 39: Effect size estimation, subgroup analysis—very low ability readers at baseline

	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
Age standardised PiRA score: very low ability readers	214 (0)	86.095 (83.983–88.208)	151 (0)	88.331 (85.336–91.327)	365 (214; 151)	-0.17 (-0.42–0.08)	0.189

As with the multilevel model for low attaining readers, the unadjusted mean for the intervention group was lower than in the control group. As shown in Table 40, the unadjusted difference in means was -2.236. In the multilevel model that accounted for the clustering of pupils in schools and also for PiRA baseline scores, geographical region, and single or multi-form entry, the observed adjusted difference in means was -2.025 age standardised scores on the PiRA. The lower bound 95% interval estimate was -5.048, while the upper bound estimate was 0.997. The point estimate and intervals

indicate that those with very low reading attainment at baseline that received PALS-UK were substantially worse off in their reading attainment as measured by the PiRA when compared to their very low attaining counterparts that were allocated to the control condition, although the data also supported some positive impact of PALS-UK for this subgroup as indicated by the upper bound estimate. The point estimate of the average treatment effect equivalises to an effect size of -0.17, or two months less progress in reading attainment among those in the PALS-UK group (Table 39). The lower bound effect size estimate here was -0.42, while the upper bound estimate was 0.08. However, the p-value here (0.189) suggests that the probability that these results were observed due to statistical uncertainty was quite high.

Finally, Table 40 and Table 41 present the results of the multilevel model and effect size estimation for the effect of PALS-UK on pupils with very high reading ability as measured by the PiRA test at baseline. At endline, there were 355 pupils scoring in the top quartile for reading attainment at baseline as measured by the PiRA: 22.94% of the primary analytical sample were categorised as having very high reading attainment at baseline in intervention schools; in control schools it was 25.45%.

Table 40: Subgroup analysis results—very high ability readers at baseline

			Intervention group		Control group		
Outcome	Unadjusted differences in means (I-C)	Adjusted differences in means (95% CI)	n (missing)	Variance of outcome	n (missing)	Variance of outcome	Pooled variance
Age standardised PiRA score: very high ability readers	-1.153	-0.380 (-3.423–2.663)	195 (0)	144.822	155 (0)	120.550	134.193

Table 41: Effect size estimation, subgroup analysis—very high ability readers at baseline

	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
Age standardised PiRA score: very high ability readers	195 (0)	110.271 (108.077–112.465)	155 (0)	111.424 (109.369–113.479)	350 (195; 155)	-0.03 (-0.30–0.23)	0.807

As with the multilevel model for high attaining readers, the unadjusted mean for the intervention group was lower than in the control group. As shown in Table 40, the unadjusted difference in means was -1.153.

In the multilevel model that accounted for the clustering of pupils in schools and also accounted for PiRA baseline scores, geographical region and single or multi-form entry, the observed difference in means was not statistically significant, with p = 0.807. The adjusted difference in means obtained from the multilevel model was -0.380. The effect size associated with this adjusted difference in means was -0.03, equivalent to zero months' additional progress in reading in the PALS-UK group (Table 41).

The conclusion of this part of the subgroup analysis is, therefore, that there was no statistically significant impact of PALS-UK on very high attaining readers. The lower bound 95% interval estimate here was -3.423 age standardised scores, while the upper bound estimate was 2.663. The point estimate of the average treatment effect equivalises to an effect size of -0.03, or zero months' additional progress in the reading attainment. The data supported a range of

scenarios, with the lower bound effect size estimate at -0.30, while the upper bound estimate was 0.23. However, the high p-value (0.807) suggests that the probability that these results were observed due to statistical uncertainty was high.

Overall, then, the results in this part of the subgroup analysis suggest no conclusive evidence of any positive or negative impact of PALS-UK among pupil subgroups with varying levels of reading attainment at baseline due to the high level of statistical uncertainty around these results.

Implementation and process evaluation

In the following sections we discuss the findings in relation to implementation compliance, implementation fidelity, perceived impact, and usual practice. Compliance deals with the question, 'Did schools complete the PALS-UK activities?' Fidelity considers the question, 'Did schools complete PALS-UK activities *as originally intended* (to the appropriate pupils and in the way they were supposed to)?' The subsequent section on outcomes discusses the extent to which PALS activities, as delivered in the trial, led to specific implementation outcomes (that is, pupils' improved reading attainment, reading fluency, and comprehension or reading confidence). Finally, the Usual Practice section looks at what was happening before PALS was implemented or after the programme ended (at baseline, and COVID-19 phase) and what was being delivered in control schools (at endline, and COVID-19 phase).

It should be highlighted here that the findings reported in this section draw largely on insights drawn from analysis of questionnaire surveys, interviews, observations of training, and peer observation logs mainly based on respondents' self-reports and perceptions. Therefore, these may not be representative of the study sample as a whole. While the same question was often asked over time in multiple surveys, it was not possible to track individual respondents across time given that cross-survey tracking information was not collected from respondents.

Compliance

Box 1: Compliance key findings

Responses by Year 5 teachers and headteachers to their respective endline surveys and data from interviews with selected case study schools shows that:

1. Attendance at initial and top-up teacher training sessions was high, both in terms of self-report (from surveys) and in terms of information provided by the delivery team. The register of attendance by the delivery team shows perfect compliance with the training. In terms of survey responses, over 95% of the teacher respondents attended the initial and top-up training.
2. Evidence from the surveys indicates perfect compliance in terms the four weeks of pupil training.
3. Reports on whether PALS-UK was delivered for 16 consecutive weeks indicate high compliance based on survey responses from headteachers, but information was missing for approximately 50% of the schools.

Crucial to successful delivery are the non-negotiables of PALS-UK: attendance at the initial training, attendance at the top-up training, and delivery of PALS-UK sessions with high fidelity. As discussed in the Impact evaluation section, compliance with the PALS-UK programme was measured as a combination of attendance at training sessions and the number weeks PALS-UK sessions were delivered to pupils in schools. This section describes these three elements in further detail. For more information please see the Analysis in the presence of non-compliance section.

Specifically, we discuss three factors related to compliance: (1) attendance at the training offered to teachers implementing PALS-UK, (2) the completion of the four week pupils training, and (3) the delivery of PALS-UK for 16 consecutive weeks.

Attendance at training sessions and use of ongoing support while delivering PALS-UK

Year 5 teachers were expected to attend an initial full day of training and a half-day of top-up training. Attendance at the training sessions was recorded by the delivery team. Based on that record, compliance was perfect because the delivery team did not let any school deliver the programme without the mandatory training. There was only one school that did not attend a top-up training event and instead training was delivered one to one on the school site.

Survey data, even though incomplete, provides useful insight on attendance at the initial and top-up training. Responses to the teacher and TA endline surveys reveal that attendance was high overall. When we asked teachers in the first endline survey if they were able to attend the initial full day training session, the majority, 94.9% (37 of 39), confirmed that they attended while 5.1% (2 of 39) indicated that they were not able. One of the teachers could not attend because they had come into the role later that year, while the other teacher confirmed that had received the training later in the year, in October 2020. A similar situation is observed in the case of the top-up training: survey respondents who indicated

they had attended the half-day top-up training were also in the majority: 94.9% (37 of 39) while 5.1% (2 of 39) indicated they had not attended this training. One teacher elaborated that they had schedule conflict, while the other indicated that their Year 5 class was not involved in the evaluation. It is important to note that all teachers who responded to the survey attended *either* an initial or top-up training session; the two teachers that reported not attending the initial training subsequently attended the top-up training while the two that reported not attending the top-up training had already attended the initial training. Hence, it can also be concluded that compliance to training attendance was overall high for Year 5 teachers based on endline survey responses.

Completion of the four weeks of pupil training

A core component of PALS-UK is the delivery of four weeks of pupil training. Survey data indicates perfect compliance. A small number of schools indicated that they even trained pupils for more than four weeks (12.8%, 5 of 39) while information was missing for five schools (11.36%) allocated to the treatment condition. The present study did not keep record of the number of sessions delivered each week, or on the number of pupils that attended each training session. The effect of absences on delivery is discussed in more detail in the Implementation fidelity section.

Completion of 16 consecutive weeks of PALS-UK

The present study did not have in place practices to consistently externally record the number of PALS-UK sessions completed and the number of consecutive weeks PALS-UK was being delivered. As discussed in the Analysis in the presence of non-compliance, we assess the number of weeks PALS-UK was delivered based on information collected in the headteacher endline survey and information from the observation checklist supplied by the delivery team. Based on the information collected from headteachers, all schools that responded to the surveys delivered 16 weeks of PALS-UK (100.0%, 23 of 23) but the information was missing for nearly half of the schools (47.7%, 21 of 44). Among these 21, it is unclear whether or not they completed the delivery.

Implementation fidelity

Box 2: Implementation fidelity key findings

Responses by headteacher and teachers to surveys, data from interviews with headteachers and teachers from selected case study schools, and data from peer observer and RA observations show that:

1. Responses to the baseline survey indicate that information received about PALS-UK prior to implementation was clear to a lesser or greater extent from the outset.
2. As reported at baseline and endline in the surveys, the majority of schools felt that they were prepared to implement PALS-UK, although there was some variation in views on this at baseline in particular.
3. Overall, PALS-UK training was deemed high quality and useful for delivering the programme by the majority of teachers who participated in the data collection activities. However, there was slightly less consensus around the usefulness of pupil training for informing pupils how to work well in pairs, as well as the usefulness of the top-up training for informing the selection of appropriate reading books for pairs.
4. Evidence from the surveys, observations, and interviews suggests that the PALS-UK programme was delivered with high fidelity. However, it is important to take into account the low survey response rates and thus the potential for selection bias when interpreting the results.
5. Generally, the PALS-UK resources were deemed useful for programme delivery. However, while the PALS-UK manual and video resources were particularly useful, evidence from the IPE indicates that survey tasks and email reminders were not deemed as useful by schools for implementation.
6. Key barriers to effective implementation included limited capacity (often related to only having a limited number of staff members trained in PALS-UK) and time to deliver sessions (often framed

within the context of the COVID-19 pandemic), competing priorities in the school, and student absenteeism (also often framed in relation to the COVID-19 pandemic). Key facilitators included the training activities, PALS-UK resources, and ongoing support.

7. Evidence from surveys and interviews indicate that schools were satisfied with ongoing and additional support provided by the delivery team. School staff also indicated that there was wider support for PALS-UK in the school and other staff members were also engaged with the process.

8. The majority of headteachers and teachers indicated in surveys and interviews that they plan to (or would like to at least) continue delivering PALS-UK beyond the end of the trial. However, it was common for schools to indicate that they could only do so if certain conditions were met, for example, reducing the number of sessions per week.

To understand better how PALS-UK was delivered in practice, we identified several elements where fidelity would be of key importance. When discussing these elements we focus on quality, content, appropriateness (for example, of training or pupil pairing), ease of delivery (enablers and barriers), and the presence of any adaptations or deviations from the original design of the intervention (in addition to what is captured under the Compliance section). The elements discussed using these lenses are the PALS-UK training (including the initial training, the top-up training, and the four-week training for pupils), pupil pairing, barriers and facilitators to delivery, the PALS-UK resources, and the support that staff received to deliver PALS-UK. Finally, we will reflect on schools' perceptions on whether they planned to continue delivering PALS-UK beyond the conclusion of the trial. These elements were selected based on the intervention's logic model and are discussed in turn below.

PALS-UK set-up and preparedness

Motivation for joining the PALS-UK programme

Headteachers in intervention schools were asked at baseline to describe their school's motivation for joining the PALS-UK programme. The most widely mentioned was that the improvement of reading and comprehension skills was a priority in the school (mentioned by 72.2%, 26 of 36). In some cases, the school's motivation was to improve these skills among specific pupil subgroups pupils eligible for the Pupil Premium, male pupils, or EAL pupils. Another widely mentioned motivation was the opportunity to participate in a research-led approach to the improvement of reading skills (mentioned by 22.2%, 8 of 36). Schools tended to mention this within the wider context of either not being sure which approach or programme to adopt for this purpose, or having tried other approaches or languages previously, but without success. Linked to this, a few headteachers also appreciated the opportunity to work with, and participate in, EEF-commissioned research (mentioned by 19.4%, 7 of 36).

Other less-often mentioned motivations to participate in the study included reading attainment being of concern in the school (13.9%, 5 of 36) and the potential for PALS-UK to have an impact on pupils (8.3%, 3 of 36).

On the other hand, one headteacher (INT-HTS3-1), in the case study interview at endline, said that staff had voiced concerns about taking part in the programme; namely, that if the programme did not work, it would present some risks given that pupils would be so close to finishing primary school:

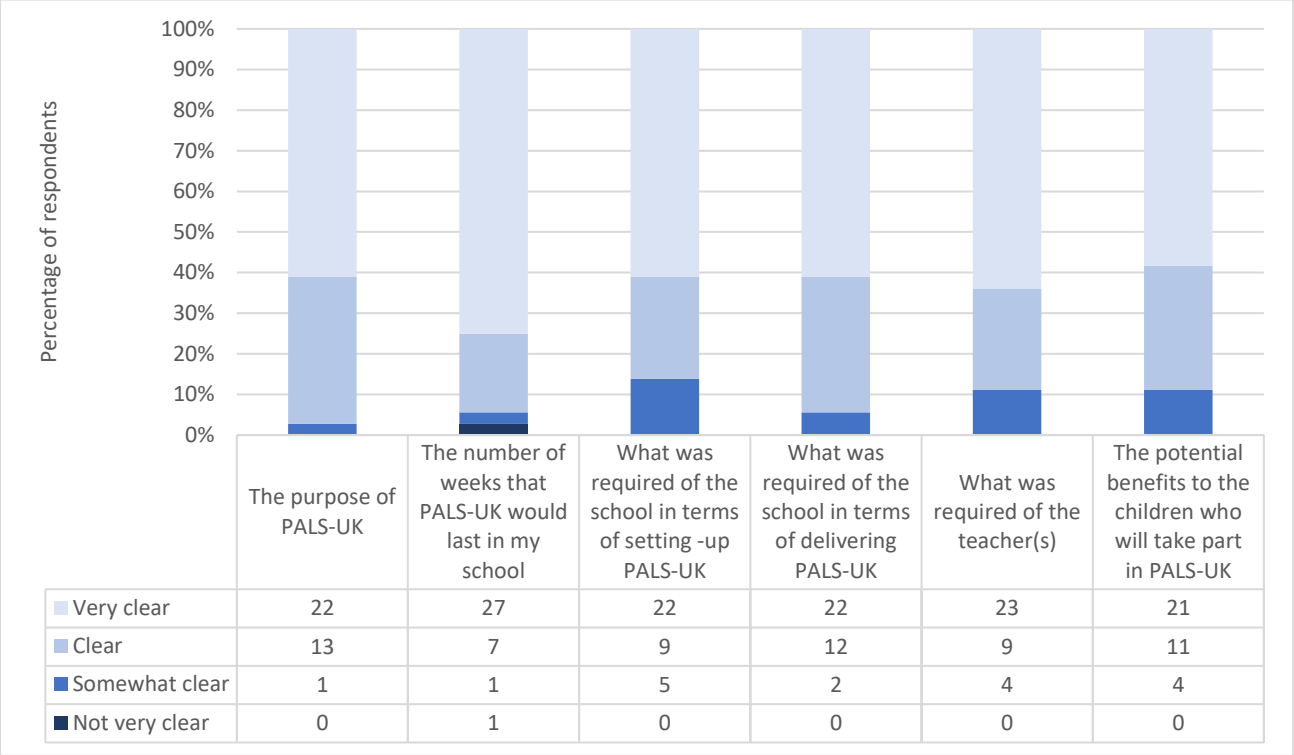
> *'[Staff were] also worried that if it didn't work it's risky because Year 5 is so close to the end of school' (INT-HTS3-1).*

Information about PALS-UK at set-up stage

At baseline, headteachers in intervention schools were asked about the clarity of the information that they received about PALS-UK prior to implementation (Figure 6). Overall, responses to the baseline survey indicate that this information was clear from the outset. Indeed, the majority of headteachers thought that the following was made clear or very clear prior to implementation: the purpose of PALS-UK (97.2%, 35 of 36); the number of weeks the programme would last (94.4%, 34 of 36); what was required of the school in terms of setting up (86.1%, 31 of 36) and delivering (94.4%, 34 of 36) PALS-UK; what was required of the teachers with regards to PALS-UK (88.9%, 32 of 36); and the potential benefits of PALS-UK for the children taking part (88.9%, 32 of 36).

A minority of headteachers indicated that the requirements of the school, the teacher(s), and the potential benefits of the programme for children were 'somewhat clear', while just one headteacher across the items indicated that something was 'not very clear', that is, the number of weeks that the programme would last.

Figure 6: Headteachers' views on the information provided about the PALS-UK programme at set-up stage



	The purpose of PALS-UK	The number of weeks that PALS-UK would last in my school	What was required of the school in terms of setting -up PALS-UK	What was required of the school in terms of delivering PALS-UK	What was required of the teacher(s)	The potential benefits to the children who will take part in PALS-UK
Very clear	22	27	22	22	23	21
Clear	13	7	9	12	9	11
Somewhat clear	1	1	5	2	4	4
Not very clear	0	1	0	0	0	0

Headteacher baseline survey question 26, n = 36 across all items. Question: 'Based on the information your school received before the start of PALS, please indicate how clear the following details were to you.'

Levels of preparedness to implement PALS-UK

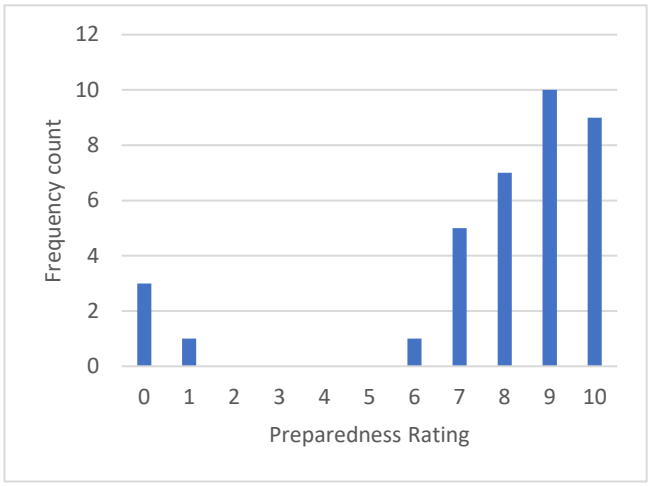
Linked to the quality of the training, headteachers in intervention schools were asked to report their school's level of preparedness for implementing PALS-UK at baseline and endline. As shown in Figure 7, views at baseline were somewhat mixed. While the majority indicated that their school was well prepared (72.2%, 26 of 36 headteachers provided a score of 8 out of 10 or above), others indicated that they were somewhat prepared (16.7%, 6 of 36 provided a score of 6 or 7 out of 10) while a few indicated that they were only prepared to a very small extent or not prepared at all (11.1%, 4 of 36 provided a score of 0 or 1 out of 10).

At endline, headteachers were asked whether their preparedness for implementing PALS-UK was appropriate. The level of unanimity was higher at this timepoint with the vast majority indicating their school's level of preparedness was appropriate (95.7%, 22 of 23) and just one indicating that this was not the case.

Overall, then, in terms of preparedness to deliver PALS-UK the majority of headteachers at baseline and endline indicated that their school was prepared, although some schools at baseline and one school at endline indicated that this was not the case.

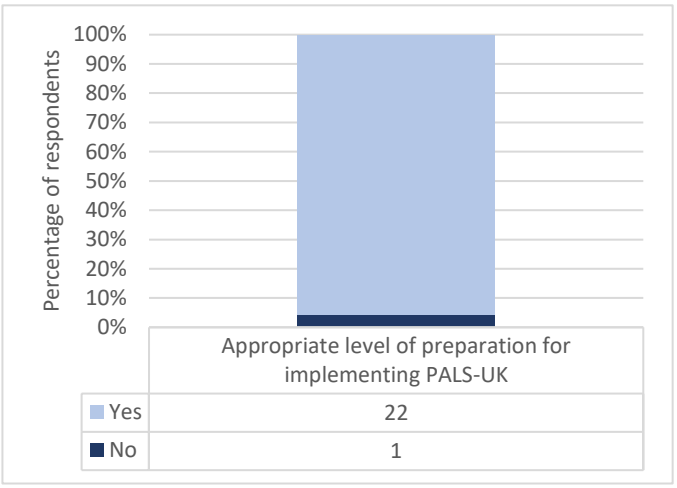
Figure 7: Headteachers' perceptions of the school's preparedness to implement PALS-UK at baseline and endline

Figure 7a Baseline



Headteacher baseline survey question 27, n = 36. Question: 'How would you rate your school's overall preparedness for implementing PALS-UK? Please rate your preparedness from 0 to 10 where 0 is completely unprepared and 10 is fully prepared.'

Figure 7b Endline



Headteacher endline survey question 21, n = 23. Question: 'Was your level of preparation for implementing PALS-UK appropriate?'

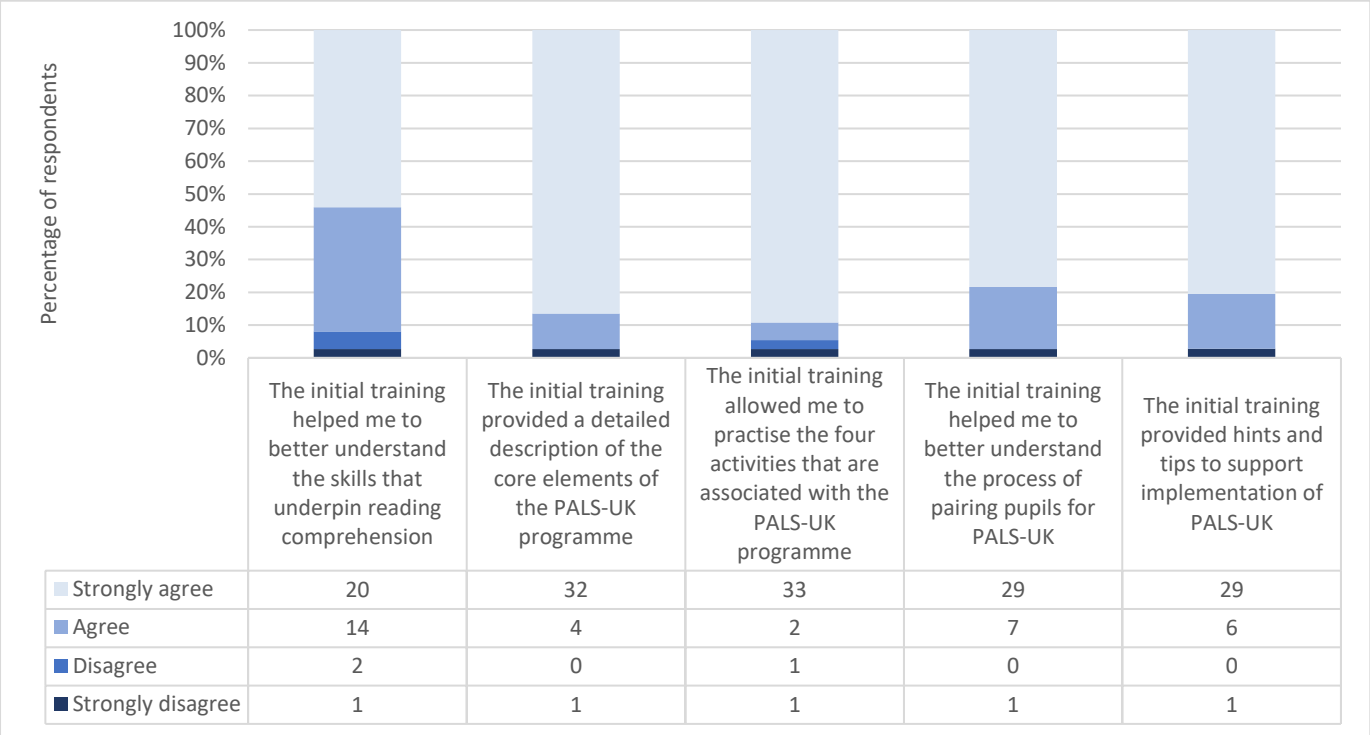
Perceived training quality

Perceptions of PALS-UK training

As described in the Intervention section, there were three elements to the PALS-UK training: (1) an initial one-day of training for teachers, (2) four-week pupil training, and (3) a half-day of top-up training for teachers following the four-week pupil training. Perceptions of each of these aspects of the training will be discussed here.

Starting first with the initial training day, Figure 8 provides an insight from the endline survey of teachers on their perceptions of its quality. Overall, findings indicate that teachers found the initial training useful for understanding and eventually delivering the PALS-UK programme. In terms of understanding, the vast majority of respondents agreed or strongly agreed that the initial training helped them to better understand the skills that underpin reading comprehension (91.9%, 34 of 37) and provided a detailed description of the core elements of the PALS-UK programme (97.3%, 36 of 37). Teachers found the initial training to be similarly useful for the more practical aspects of delivering PALS-UK: the vast majority indicated that it allowed them to practice the four activities associated with PALS-UK (94.6%, 35 of 37), better understand the process of pairing pupils (97.3%, 36 of 37), and provided hints and tips to support the overall implementation of PALS-UK (97.2%, 35 of 36).

Figure 8: Teachers' perceptions of the initial PALS-UK training



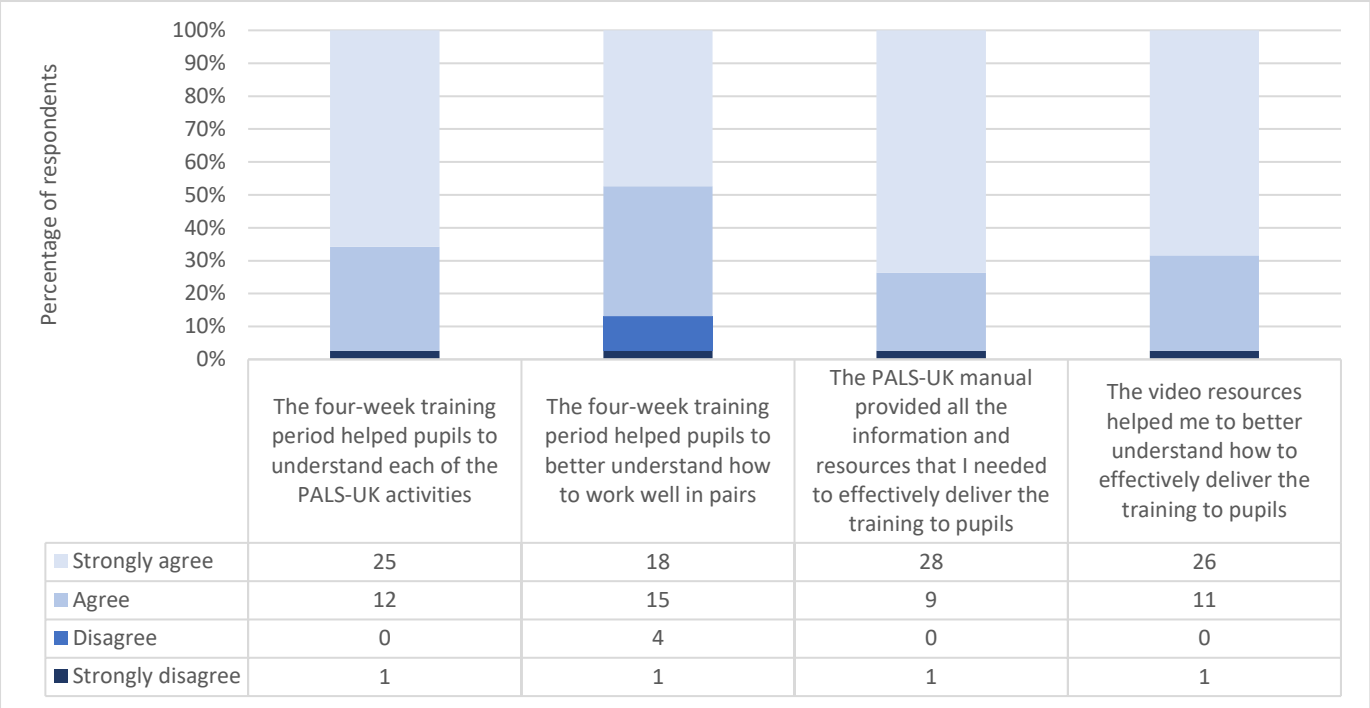
	The initial training helped me to better understand the skills that underpin reading comprehension	The initial training provided a detailed description of the core elements of the PALS-UK programme	The initial training allowed me to practise the four activities that are associated with the PALS-UK programme	The initial training helped me to better understand the process of pairing pupils for PALS-UK	The initial training provided hints and tips to support implementation of PALS-UK
Strongly agree	20	32	33	29	29
Agree	14	4	2	7	6
Disagree	2	0	1	0	0
Strongly disagree	1	1	1	1	1

Teacher endline survey question 5, n = 37 for items 1–4 (from left to right); n = 36 for item 5. 'Thinking about the initial full day training session, please specify the extent to which you agree with the following statements.'

Overall, findings from the survey indicate that teachers perceived the initial training to be useful for understanding and delivering the programme.

Focusing next on the four-week training of pupils in the PALS-UK approach, initial insights from the endline survey of teachers in interventions schools indicated that, overall, teachers felt well prepared to deliver this training and that it was useful to the pupils that received it (Figure 9). The vast majority of teachers agreed or strongly agreed that the manual provided sufficient information to effectively deliver this training (97.4%, 37 of 38) while the same proportion agreed that the video resources had also been useful (97.4%, 37 of 38).

Figure 9: Teachers' perceptions of the four-week training of pupils in the PALS-UK approach



	The four-week training period helped pupils to understand each of the PALS-UK activities	The four-week training period helped pupils to better understand how to work well in pairs	The PALS-UK manual provided all the information and resources that I needed to effectively deliver the training to pupils	The video resources helped me to better understand how to effectively deliver the training to pupils
Strongly agree	25	18	28	26
Agree	12	15	9	11
Disagree	0	4	0	0
Strongly disagree	1	1	1	1

Teacher endline survey question 10, n = 38 across all four items. 'We would like to know more about your views on the four-week training for pupils as part of PALS-UK. Please indicate the extent to which you disagree or agree with the following statements.'

With regards to the usefulness of the training for pupils, insights from the endline survey of teachers revealed a similar positive view; the vast majority of teachers agreed or strongly agreed that the four-week training of pupils had helped them to understand each of the four PALS-UK activities (97.4%, 37 of 38) and how to work well in pairs (86.8%, 33 of 38). It is important to note here that a slightly higher proportion of respondents disagreed or strongly disagreed with the statement that the training had helped pupils understand how to work well in pairs (13.2%, 5 of 38), although this still represented only a small minority of the respondents. Nonetheless, it may suggest that there were some issues with pairing pupils in ways that ensured that they were able to work well together on the PALS-UK activities.

Overall, insights from the endline survey of intervention teachers indicated that the four-week training of pupils was a positive experience, both in terms of how teachers were prepared to deliver this training and the usefulness of the training for pupils in terms of their participation in the PALS-UK programme.

Looking finally at the half-day top-up training that took place following the delivery of the four-week pupil training, insights from the endline survey revealed that teachers who attended were generally in agreement that this training was useful (Figure 10). This sentiment was also evidenced in the case study interviews:

> *'Really good and clear, and the manual we were given really helped. The refresher was also really useful and made me feel confident to teach it. The videos were also really helpful' (INT-TS2-1).*

> *'Yes, the step by step training [sessions] were really useful. The refresher was great to brush up on things you might have forgotten so I'd say the refresher was most beneficial' (INT-TS3-1).*

The majority of teachers agreed or strongly agreed that the top-up training was useful for understanding how to select appropriate books for pupil pairs (83.8%, 31 of 37), however, a minority disagreed or strongly disagreed with this (16.2%, 6 of 37), perhaps indicating that the selection of books for pupils was problematic for some intervention schools. Nonetheless, the vast majority of teachers were also in agreement that the top-up training was a good opportunity to share good practices related to delivering PALS-UK with other teachers (91.9%, 34 of 37). Again, this view was expressed in the case studies also:

> *'Oh, it was brilliant. They went through how children learn to read such as their language acquisition and then spoke through the implementation of PALS. And we got to meet other people and talked about what PALS is*

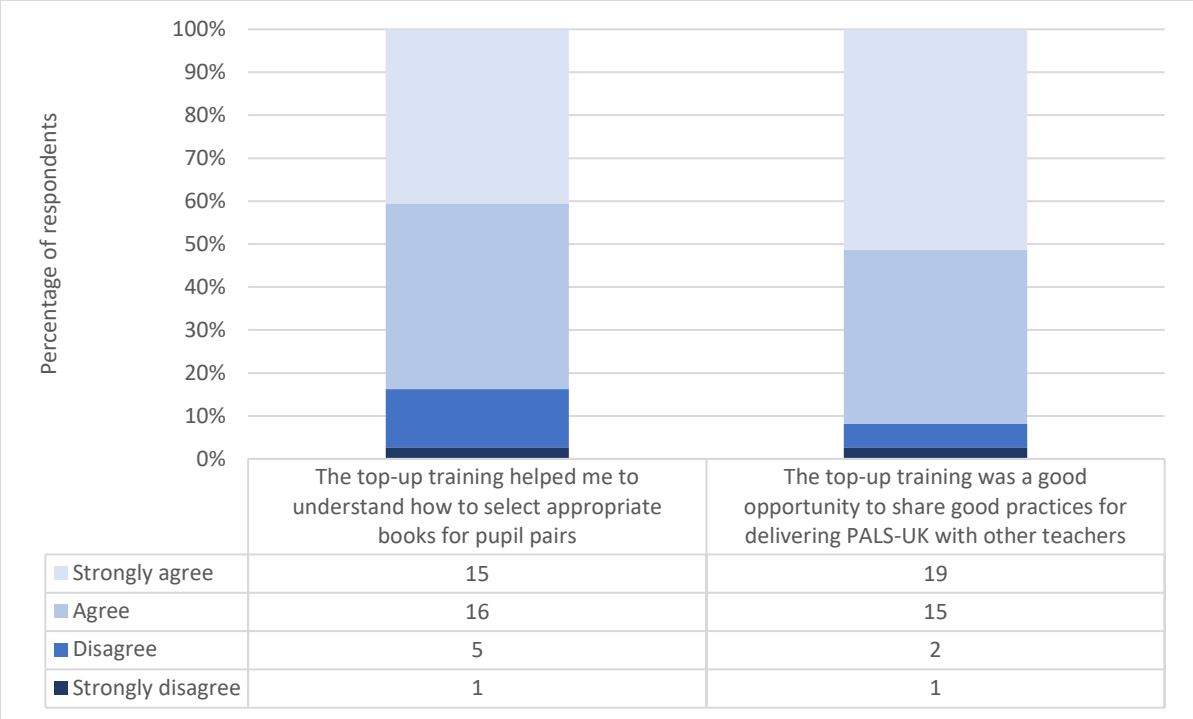
like and we could practice it before we went home so that was really helpful to see what we'd have to teach' (INT-TS1-1).

Another interviewee mentioned in a case study interview that the opportunity to interact with others in the training made the process more engaging:

'It was really, really well presented and engaging. It was really interactive, I like that you had a chance to physically practice with each other and ask questions' (INT-POS2-1).

Overall, then, the majority of teachers responding to the endline survey were in agreement that the top-up training was useful for delivering PALS-UK.

Figure 10: Teachers' perceptions of the PALS-UK half-day top-up training



	The top-up training helped me to understand how to select appropriate books for pupil pairs	The top-up training was a good opportunity to share good practices for delivering PALS-UK with other teachers
Strongly agree	15	19
Agree	16	15
Disagree	5	2
Strongly disagree	1	1

Teacher endline survey question 8, n = 37 across both items. 'Thinking about the half-day top-up training, please specify the extent to which you agree with the following statements.'

When asked whether there was anything that could be done to improve the training, one case study interviewee noted that having another top-up session would have been useful (INT-TS3-1). Open text responses from the endline surveys included the following suggestions: a deeper explanation and modelling of the triad approach for odd numbers of children; facilities to enable real time conversations about successes and issues in the classroom related to PALS-UK; more constructive advice and critique during school visits; provision of a video example of lower-attaining readers participating in the programme; spend more time practising session delivery and less time presenting the research; and reducing the length of the overall training.[18]

Overall, the main message emerging from the IPE data around training for PALS-UK was that it was deemed high quality and useful for delivering the programme by the majority of teachers who participated in the data collection activities. A small minority of school staff did not perceive that the pupil training was useful for informing pupils how to work well in pairs, and that the top-up training was useful for informing the selection of appropriate reading books for pairs.

[18] As these insights were gained from open text responses, the evaluation team are unable to provide further context to these suggestions.

Programme delivery

In this section we will gather relevant insights from the IPE data collection activities to explore the extent to which activities related to the programme were delivered with high fidelity. We will also analyse barriers and facilitators to effective implementation as reported by school staff. Within this, we will assess the extent to which the PALS-UK resources (including the PALS-UK manual) were useful when delivering the intervention.

Implementation of PALS-UK activities—pupil training, pairing, and structured reading in pairs

As outlined in the Intervention section, a number of core activities should have taken place in intervention schools to implement the PALS-UK approach with high fidelity. Key aspects of the PALS-UK logic model were that (1) pupils were trained in the PALS-UK approach for four weeks, (2) children were paired based on their reading ability (see Intervention section for more detailed information on the guidelines for pairing), (3) pupil pairs were swapped every four weeks, and (4) children partook in paired structured reading three times a week for the duration of the programme (16 weeks). The extent to which these activities took place will be explored in some detail here.

Headteachers in intervention schools were asked whether PALS-UK delivery activities had been implemented by the time of the endline survey in their school (Figure 11). While the response rate among headteachers to the endline survey was relatively low (n = 23), all headteachers that responded indicated that by endline:

- Year 5 teachers had spent four weeks training pupils in the PALS-UK approach;
- Year 5 pupils had been paired by teachers to take part in the reading activities; and
- Year 5 pupils had worked on reading in pairs three times a week for 16 weeks during structured reading activities.

However, it is important to take into account the low response rate and thus the potential for selection bias; indeed, it is possible that schools replying to the endline surveys are among those more engaged in the programme and more likely to have carried out programme activities. Equally, it is possible that schools had completed the delivery of PALS-UK but found it difficult to respond to the evaluation survey.

While these survey findings indicate that key PALS-UK activities had taken place, case study insights suggest that this was not without issues. More than one interviewee at endline indicated that scheduling sessions could be problematic:

> *'Resourcing was fine because they were able to get everything to us that we needed. Timing-wise, it got a bit tricky around Christmas time when we'd have tests and also trying to fit it around normal reading time. There was maybe an odd week where we missed a session but mostly we tried to fit it in because I know PALS wanted us to stick to the same times every day' (TS1-1).*

Nonetheless, there was evidence from the case studies that the high levels of implementation fidelity documented in the surveys translated to effective implementation among the children themselves:

> *'Yeah, definitely they were really engaged with it, which is something I really enjoyed about PALS. The children took a more active approach than in a traditional guided reading activity and were all very switched on. They definitely got a lot out of it and [were] picking up [that] other children were making errors. They followed procedures really well … it was a well-oiled machine' (T2-1).*

> *'Yes, it was quite funny because then they'd begin correcting each other in long haul lessons or even correcting me!' (TS1-1).*

However, some interviewees noted that PALS-UK sessions were not always implemented as correctly as they could be, were quite long, and that some parts of the approach were more useful than others:

> *'At the beginning I think we needed that time to become more experienced with it, but I think some children found it repetitive towards the end. The sessions were quite long, around 40 minutes, but since we saw a good impact on fluency scores, I think it was worth it' (TS1-1).*

> *'I think the children found it tricky with paragraph shrinking and making sure that what the other child was saying was right, because sometimes they'd just skip that part. I think fluency was the best aspect and correcting one another when they made mistakes' (TS1-1).*

Furthermore, the vast majority of these headteachers reported that the Year 5 teachers had changed pairs for structured reading approximately every four weeks (91.3%, 21 of 23) and Year 5 pupils with SEND who required additional support had been paired with a TA for structured reading activities if deemed necessary (95.7%, 22 of 23). Case study interviews yielded more insights on the pairing of pupils. First, it was most common for interviewees to report that they were pairing based on reading ability:

> *'Me and my partner ordered all of the standardised scores for reading and split it down the middle. We paired the highest scoring person in the class with the highest of the lowest performing group' (T2-1).*

However, other interviewees also noted that they were pairing on other factors too, such as how well the pupils were getting along:

> *'We have two Year 5 classes, lower-middle in one and middle-highers in another so that there were 20 in each class. Some children don't work well together so we had to select pupils who would get on with it and work well together' (TS2-1).*

Some interviewees also mentioned that pupil pairs were switched every four weeks as per the logic model:

> *'Every four weeks because pupils like to swap partners' (INT-TS2-1).*

However, in a few instances, deviations to the pairing approach were reported by school staff, for example, due to pairs having difficulties working together:

> *'[Pupils pairs were swapped] every two to three weeks because some children didn't enjoy working with their partners' (INT-TS3-1).*

While absences did affect the pairing of pupils, one teacher noted that this was an area where the programme could be adapted to mitigate the impact of absences:

> *'The children were able to manage with the sheet they were given, they could follow it themselves, and were able to get on even if it was a different book they were reading; they adapted to it well' (INT-TS2-1).*

As an additional robustness check on implementation fidelity, observations of PALS-UK delivery were undertaken by peer observers and research assistants from the delivery team. Observations of activities centred around five core themes were carried out: classroom set-up, partner reading with retell, paragraph shrinking, prediction relay, and classroom management. For each theme, the number and proportion of each component (or sub-activity) completed correctly was recorded. As an overall measure of implementation fidelity, unweighted and weighted percentages for the number of correctly implemented PALS-UK activities were calculated. The unweighted percentage assumes that each theme (for example, classroom set-up) holds equal weight regardless of the number of components that were observed within that theme. This is achieved by taking the average of the percentage of correctly completed components across the five themes, thus ignoring the number of components per theme and, in doing so, assigning equal weight to each theme. The weighted percentage is based purely upon dividing the number of individual components observed correctly by the total number of components observed (see the notes below Table 42 for the number of components per theme). This approach assigns greater weighting to themes with a higher number of components.

As an example of how the unweighted and weighted measures differ, the classroom set-up theme comprises six components while the prediction relay theme has 22 components. The unweighted percentage assigns equal weighting to the classroom set-up and prediction relay themes, despite the fact that the latter has 16 more components. In the weighted measure, the prediction relay theme would be assigned more weight given that it has 16 more components, and all of these are taken into account when constructing the weighted measure.

The PALS-UK delivery team, in collaboration with the programme developer, define implementation quality using this observation data as follows:

- less than 70%: poor implementation fidelity;

- 70%–80%: adequate implementation fidelity;
- 80–90%: very good implementation fidelity;
- 90% plus: excellent implementation fidelity.

The categories are informed by the U.S. version of the implementation fidelity checker.

Table 42 presents all observation data for case study schools, at the school level. The full dataset providing information for each observation undertaken in all case study schools can be found in Appendix J: Full observation data. Using the unweighted percentage as the measure, all eight case study schools, on average, were scored 90% or higher, indicating that PALS-UK was implemented with excellent fidelity.

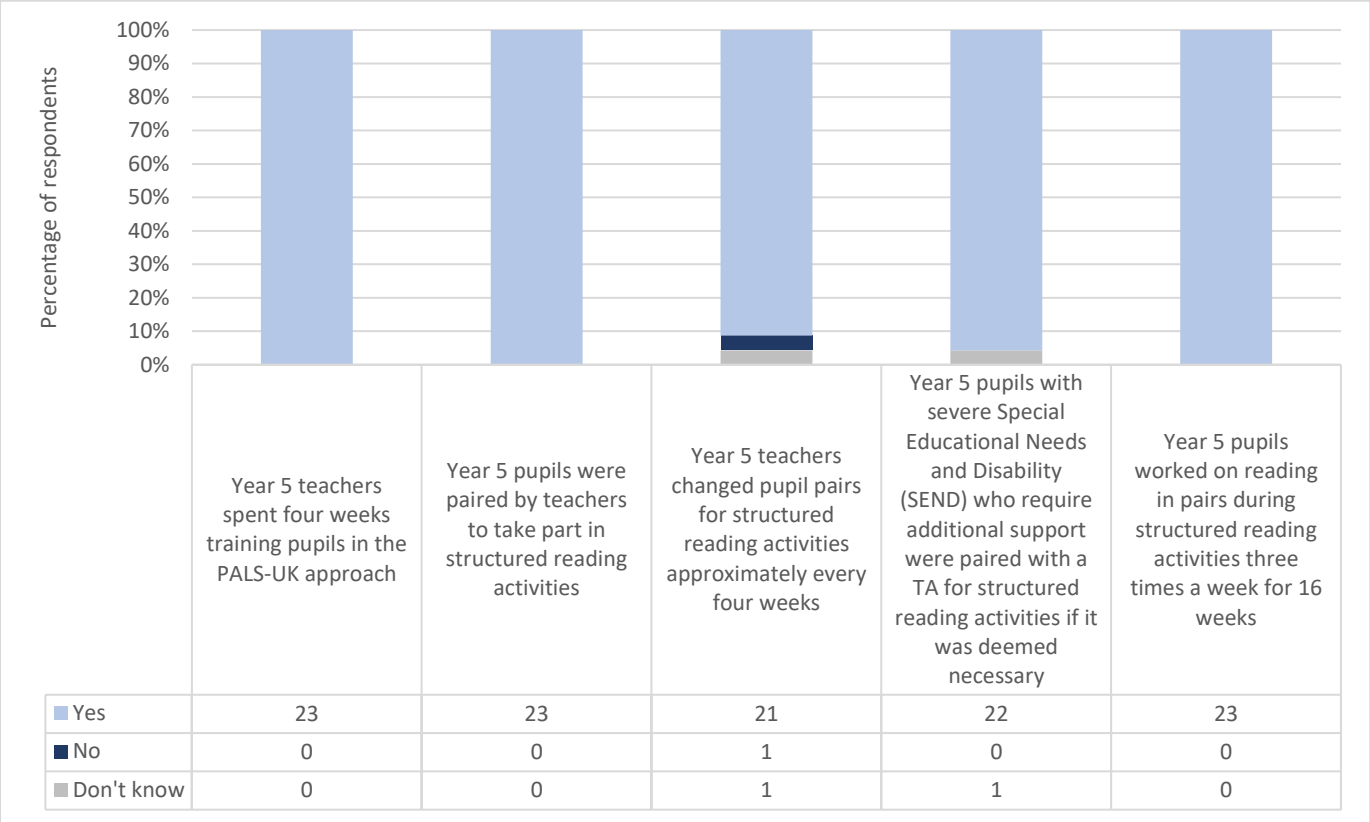
Table 42: Implementation fidelity as measured by the observation data, school level

		School average % correct activities					School average overall % correct PALS-UK implementation	
Case study ID	No. of obser-vations	Classroom set-up	Partner reading with retell	Paragraph shrinking	Prediction relay	Classroom management	Unweighted	Weighted
School 1	5	100.0%	86.8%	83.0%	92.4%	100.0%	93.2%	89.7%
School 2	5	100.0%	98.6%	97.7%	98.9%	100.0%	99.2%	98.8%
School 3	2	100.0%	82.1%	90.9%	100.0%	100.0%	95.0%	93.2%
School 4	5	100.0%	83.9%	92.4%	97.7%	100.0%	96.0%	94.8%
School 5	2	100.0%	96.4%	79.5%	100.0%	100.0%	94.6%	90.5%
School 6	3	100.0%	88.1%	98.5%	81.8%	100.0%	94.6%	75.2%
School 7	6	100.0%	91.7%	98.2%	99.1%	100.0%	97.1%	95.4%
School 8	4	100.0%	98.2%	75.8%	86.3%	100.0%	94.0%	90.2%

School level percentages have been calculated by taking the mean percentage across each of the observations undertaken in the school.

Implementation as measured by the weighted fidelity percentage yielded similarly positive results, albeit slightly more mixed. Indeed, six of the eight case study schools achieved a score of 90% or above, indicating excellent implementation fidelity, with one other case study school achieving a score just below 90% (89.7%). However, one case study school did score considerably lower on the weighted measure (75.2%), although this score still indicates adequate implementation fidelity. Overall, the key message arising from the analysis of observation data is that, on average, schools were implementing the various PALS-UK activities with excellent fidelity.

Figure 11: Headteachers' responses on delivery of PALS-UK activities at endline



	Year 5 teachers spent four weeks training pupils in the PALS-UK approach	Year 5 pupils were paired by teachers to take part in structured reading activities	Year 5 teachers changed pupil pairs for structured reading activities approximately every four weeks	Year 5 pupils with severe Special Educational Needs and Disability (SEND) who require additional support were paired with a TA for structured reading activities if it was deemed necessary	Year 5 pupils worked on reading in pairs during structured reading activities three times a week for 16 weeks
Yes	23	23	21	22	23
No	0	0	1	0	0
Don't know	0	0	1	1	0

Headteacher endline survey question 22, n = 23 across all items. 'Please indicate whether the following activities occurred during the implementation of the PALS-UK programme in your school.'

Supporting this evidence from headteachers, at endline teachers were asked whether they were also a peer observer for another teacher delivering PALS-UK in their school. While only a minority indicated that this was the case (15.4%, 6 out of 39),[19] all six peer observers indicated that they believed that the programme was being delivered as intended to a great extent based on the delivery that they had observed in their school.[20]

Overall, then, evidence from the surveys, observations, and interviews suggests that the programme had been delivered with high fidelity in schools where the headteacher had provided a response. However, as mentioned above, it is important to take into account the low response rate and thus the potential for selection bias to affect these results.

Usefulness of PALS-UK resources for implementation

In this section we will explore the extent to which PALS-UK resources were useful for programme implementation—the usefulness of the PALS-UK manual, reading books provided, survey tasks, delivery videos, email reminders to switch pupil pairs and email tips for reading book selection.

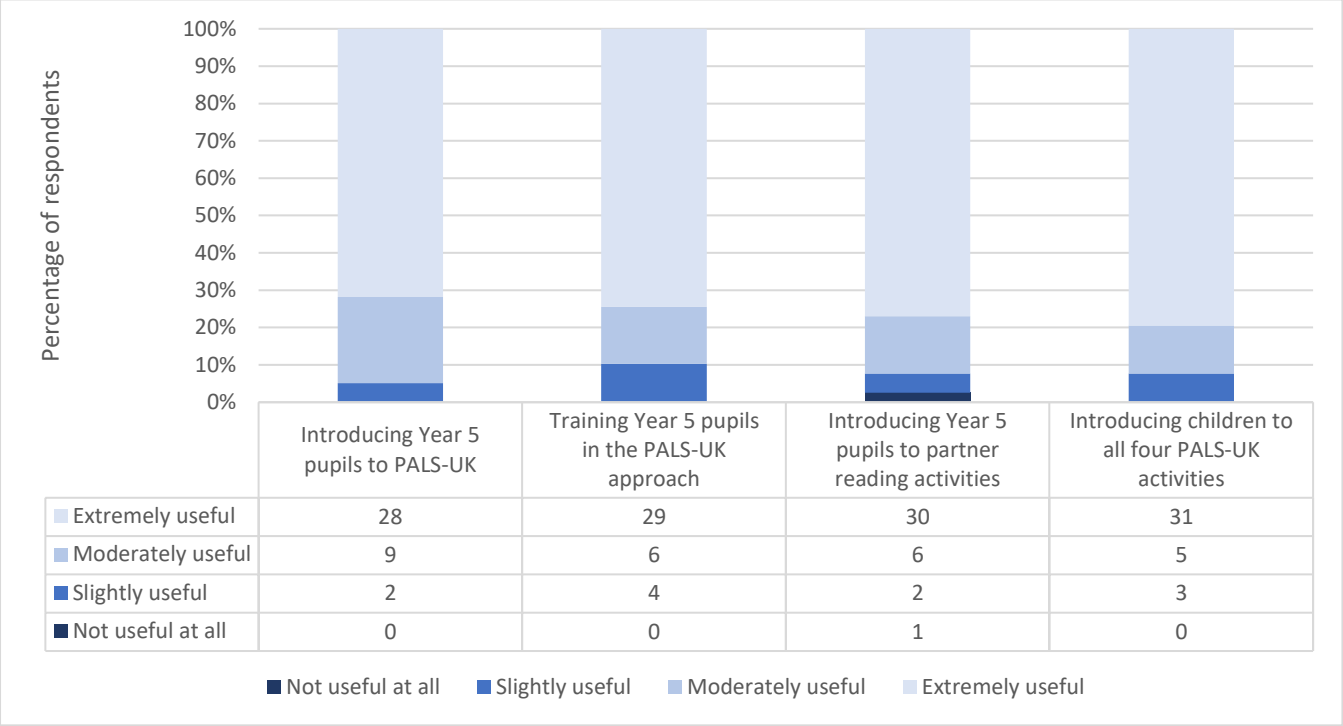
Focusing first on the PALS-UK manual, Figure 12 shows that, among the majority of teachers in intervention schools that responded to the endline survey, the manual was perceived to be useful for various aspects of programme delivery. The majority of respondents perceived the manual to be extremely useful for introducing Year 5 pupils to PALS-UK (71.8%, 28 of 39),for training Year 5 pupils in the PALS-UK approach (74.4%, 29 of 39), for introducing Year 5 pupils to partner reading activities (76.9%, 30 of 39, and for introducing children to all four PALS-UK activities (79.5%, 31 of 39). This sentiment was reflected in the case study interviews:

[19] Teacher endline survey question 26 (n = 39): 'Were you also a peer observer for another Year 5 teacher delivering the PALS-UK programme in your school?'
[20] Teacher endline survey question 27 (n = 6): 'Thinking specifically about your time as a peer observer, in your view, to what extent was the PALS-UK programme delivered as intended in the classroom(s) that you observed?'

'The manual we were given was excellent, as were other materials, throughout the whole of the trial. It's something that I and the children could refer back to' (T2-1).

Just one teacher (2.6%, 1 of 39) reported that the PALS-UK manual was not useful at all for one aspect of programme delivery; this was introducing Year 5 pupils to partner reading activities.

Figure 12: Teachers' perceptions of the usefulness of the PALS-UK manual for programme delivery



	Introducing Year 5 pupils to PALS-UK	Training Year 5 pupils in the PALS-UK approach	Introducing Year 5 pupils to partner reading activities	Introducing children to all four PALS-UK activities
Extremely useful	28	29	30	31
Moderately useful	9	6	6	5
Slightly useful	2	4	2	3
Not useful at all	0	0	1	0

Teacher endline survey question 13, n = 39 across all items. 'We would like to know about your views on the usefulness of the PALS-UK manual. Please indicate how useful (if at all) the PALS-UK manual was for the following aspects of programme delivery.'

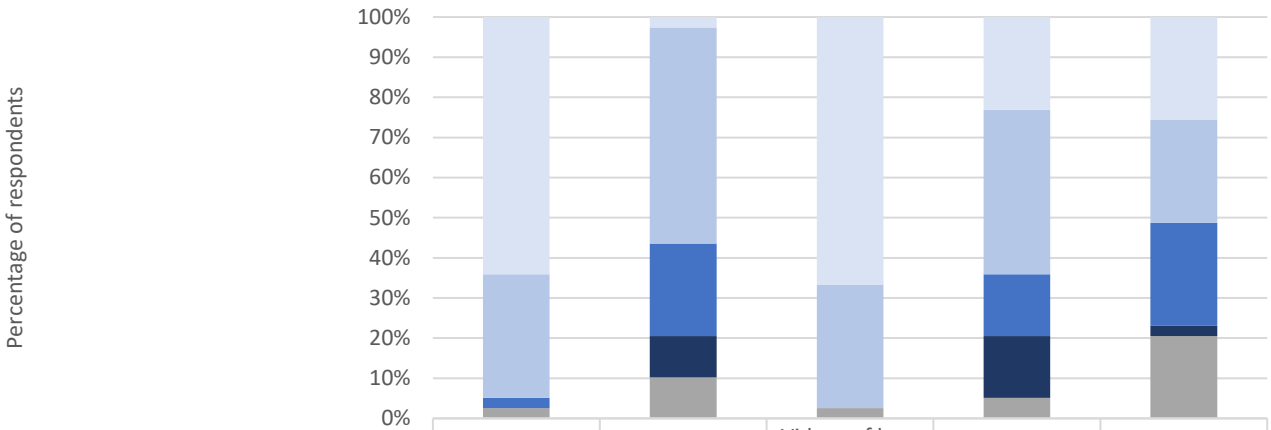
Looking at other PALS-UK resources (including reading books, survey tasks, videos, email reminders, and email tips), teachers generally deemed them useful for programme delivery although responses were more varied (Figure 13). Looking first at the selection of reading books provided for delivery, the majority of teachers found them to be extremely or moderately useful (94.9%, 37 of 39), with just one respondent indicating that they were slightly useful. These positive sentiments towards the selection of reading books was reflected in the case studies:

'They really loved them! Especially the Yeti in the Playground one! There was such a great range for different interests but also for different reading levels too' (TS1-1).

Similarly, the vast majority of teachers indicated that videos of how to deliver PALS-UK activities were deemed extremely or moderately useful for programme delivery (97.4%, 38 of 39), which once again was reflected in the case studies:

'Really useful, the manual especially, which covered mini lessons on areas where the children might be struggling. The videos on the USB were helpful too, the children could learn from them and see how PALS should look' (TS1-1).

Figure 13: Teachers' perceptions of the usefulness of other PALS-UK resources for programme delivery



	Selection of reading books	Survey tasks	Videos of how to deliver the PALS-UK activities	Emails with reminders to switch pairs	Emails with tips for book selection
Extremely useful	25	1	26	9	10
Moderately useful	12	21	12	16	10
Slightly useful	1	9	0	6	10
Not useful at all	0	4	0	6	1
N/A; did not use; did not have access to	1	4	1	2	8

Teacher endline survey question 15, n = 39 across all items. 'In your view, how useful were the following resources for delivering the PALS-UK programme?'

Looking next at the survey tasks, responses among teachers were more varied here. While the majority found them to be moderately useful (53.8%, 21 of 39), just one teacher found them to be extremely useful while almost one quarter (23.1%, 9 of 39) found them to be slightly useful and 10.3% teachers found them to be not useful at all (4 of 39).

Responses to teacher endline survey items around emails related to PALS-UK were also mixed. A slight majority found reminder emails to switch pairs extremely or moderately useful (64.1%, 25 of 39); 15% found them to be only slightly useful or not useful at all (15.4%, 6 of 39 for each response option). Finally, emails with tips for book selection were also met with a mixed response: just over one quarter of teachers found these emails to be extremely useful, moderately useful, or slightly useful (25.6%, 10 of 39 per response option). Only one teacher found them not useful at all and just over one-fifth indicated that they did not use these emails or could not access them (20.5%, 8 of 39).

Overall, insights from the surveys and interviews revealed that generally, the PALS-UK resources were deemed useful for programme delivery. However, while the manual and video resources were particularly useful, evidence from the IPE indicates that survey tasks and email reminders were not deemed as useful by schools for implementation.

Barriers and facilitators to effective implementation

In this section we will begin to explore and unpick any barriers and facilitators to effective implementation. This will draw on a combination of the IPE surveys and case studies.

First, looking at barriers, responses to the endline surveys from teachers and headteachers revealed some areas of unanimity and other areas where responses were more varied. First of all, it is important to note that among headteachers, none of the items presented below were deemed to be a major barrier to PALS-UK delivery. Nonetheless, the majority of teachers and headteachers agreed that limited capacity was not a barrier to delivery (66.7%, 26 of 39 and 65.2%, 15 of 23 reported 'not at all' respectively). However, more teachers than headteachers indicated that competing priorities in the school were 'somewhat a barrier' or 'a barrier to a great extent' (30.8%, 12 of 39 of teachers compared to 13.0%, 3 of 23 for headteachers). Interestingly, case studies seemed to suggest that a key barrier to effective implementation was finding the time to schedule PALS-UK sessions, which was often framed within the wider context of challenges posed by the COVID-19 pandemic:

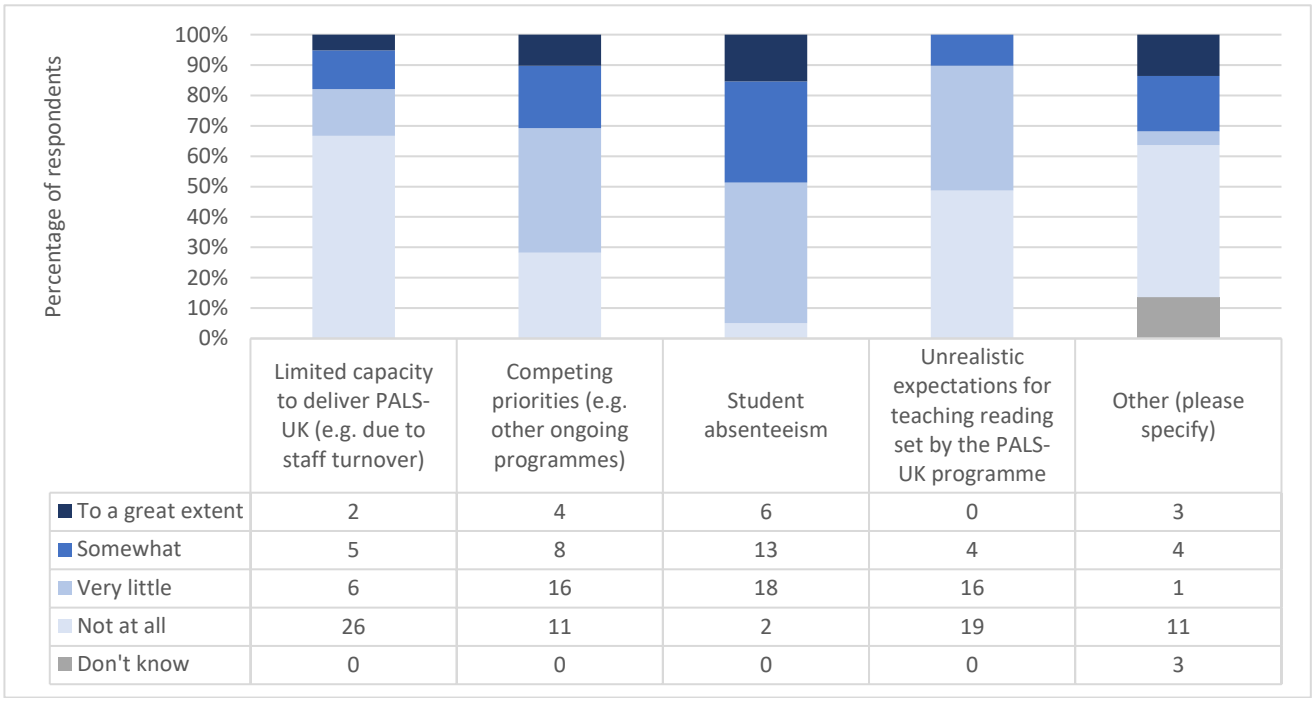
'Biggest challenge was time: they found three days a week a big investment and because of the lockdown we haven't seen the end product. So they'll probably reduce the time when we continue with' (HTS3-1).

Teachers and headteachers were also asked about whether unrealistic expectations set by the programme for teaching reading was a barrier to effective implementation; teacher and headteacher views were similar here, with the majority indicating that this was not a barrier at all, or was a barrier only to a very little extent (89.7%, 35 of 39 for teachers; 91.3%, 21 of 23 headteachers).

Teachers were also asked about whether student absenteeism was a barrier to effective implementation: almost half indicated that this was at least somewhat a barrier, if not to a great extent (48.7%, 19 of 39). Given that the PALS-UK implementation period ended (April 2020) at a similar time as the movement to remote learning as well as other effects of the COVID-19 pandemic started to manifest in the U.K. (March 2020), issues related to student absenteeism brought on by self-isolation may have affected the latter stages of PALS-UK delivery. However, it is difficult to assess from the data collected in the IPE the extent to which student absenteeism related to COVID-19 affected PALS-UK implementation in this trial.

Figure 14: Teacher and headteacher perceptions of barriers to effective implementation of PALS-UK

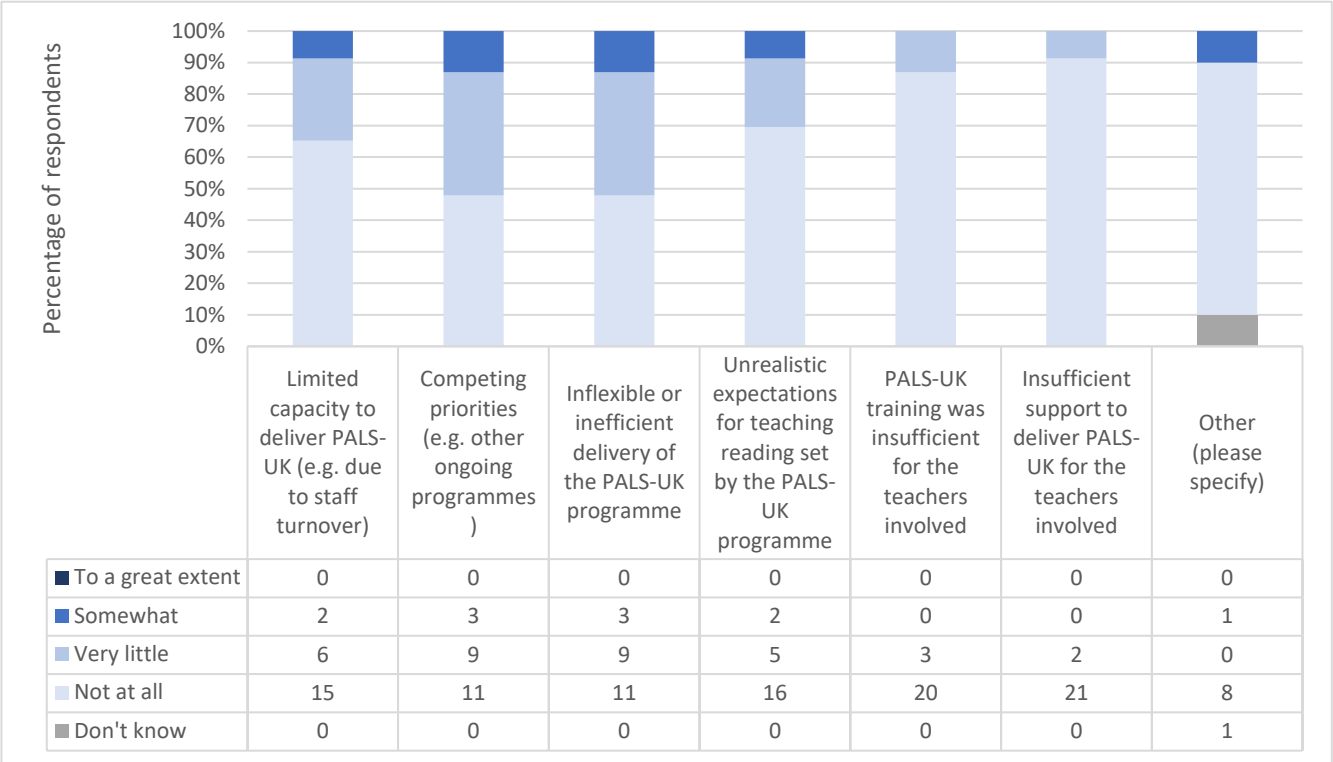
Figure 14a: The perceptions of teachers



	Limited capacity to deliver PALS-UK (e.g. due to staff turnover)	Competing priorities (e.g. other ongoing programmes)	Student absenteeism	Unrealistic expectations for teaching reading set by the PALS-UK programme	Other (please specify)
To a great extent	2	4	6	0	3
Somewhat	5	8	13	4	4
Very little	6	16	18	16	1
Not at all	26	11	2	19	11
Don't know	0	0	0	0	3

Teacher endline survey question 25: n = 39 for items 1–4 (from left to right); n = 22 for item 5. 'Please indicate the extent to which you consider the following factors to have been barriers or obstacles to the effective implementation of the PALS-UK programme.'

Figure 14b The perceptions of headteachers



	Limited capacity to deliver PALS-UK (e.g. due to staff turnover)	Competing priorities (e.g. other ongoing programmes)	Inflexible or inefficient delivery of the PALS-UK programme	Unrealistic expectations for teaching reading set by the PALS-UK programme	PALS-UK training was insufficient for the teachers involved	Insufficient support to deliver PALS-UK for the teachers involved	Other (please specify)
To a great extent	0	0	0	0	0	0	0
Somewhat	2	3	3	2	0	0	1
Very little	6	9	9	5	3	2	0
Not at all	15	11	11	16	20	21	8
Don't know	0	0	0	0	0	0	1

Headteacher endline survey question 24: n = 23 for items 1–6 (from left to right); n = 10 for item 7. 'Please indicate the extent to which you consider the following factors to have been barriers or obstacles to the effective implementation of the PALS-UK programme.'

Headteachers were also asked about whether insufficient training to deliver PALS-UK, along with insufficient support to deliver the programme were barriers to effective implementation for teachers. In both cases, the vast majority of headteachers indicated in the endline surveys that these were not barriers at all (87.0%, 20 of 23 and 91.3%, 21 of 23 respectively). Interestingly, one case study interviewee suggested that gaining staff buy-in for the programme was initially a challenge, which would seemingly contradict the overarching message from the surveys around support for PALS-UK delivery:

> *'It was around changing the mindset of staff: they were quite slow to train the children up. Challenges of getting children paired up accurately, to have the confidence to interject when their partner is reading. Enablers have got to be that you need an open attitude to guided reading and how much children can get out of it. It's got to have that senior leadership present to lead and implement it' (HTS2-1).*

Of the teachers and headteachers that indicated that there were other barriers, among those mentioned as open text survey responses included:

- not having a wider group of staff members trained in PALS-UK to account for illnesses and absences;
- the programme being too prescriptive in nature at times;
- difficulties finding appropriate pairs when swapping pupil pairs;
- the length of PALS-UK sessions;
- difficulties with children working together to participate in structured reading;
- time constraints in the school timetable;
- reading books not appropriate for some of the participating children; and

- difficulties with implementation owing to behavioural issues in the class—although this respondent indicated that the issue became less significant as they progressed through the programme.[21]

Furthermore, multiple interviewees highlighted that it was difficult at times to keep pupils engaged with the programme:

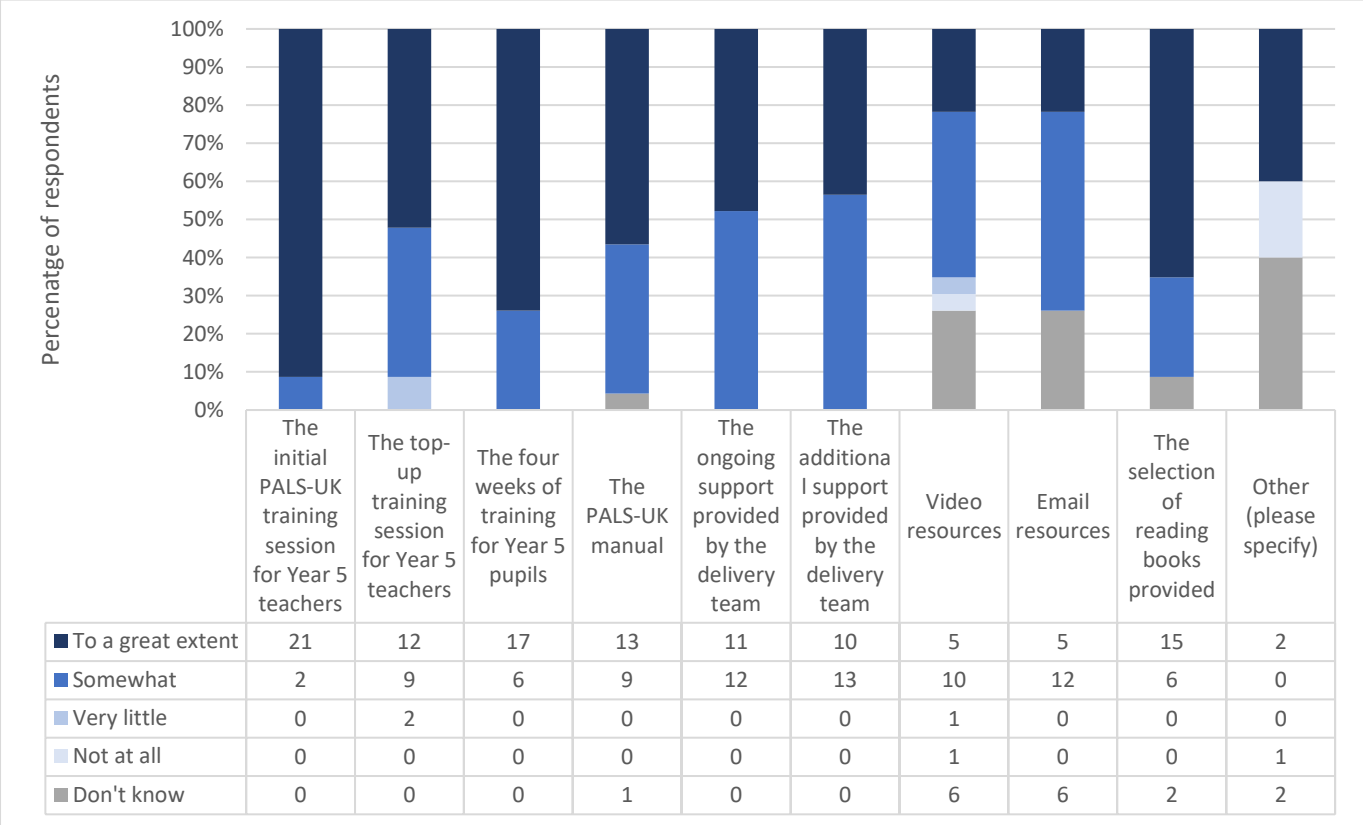
> *'That's something which came from the children: they were getting bored because it was so long and they lost interest. So it should have been shorter' (TS3-1).*

On the other hand, other schools elaborated in the case study interviews that children were highly engaged in the programme, suggesting that levels of engagement among pupils were mixed across the intervention schools:

> *'The engagement with the reading and different ideas … they'd perform better with a partner they liked. But being able to work together was really effective, it helped them develop more skills apart from reading, such as peer feedback. They got into a routine, so they knew exactly what they were doing, following the procedures really well' (TS3-1).*

Moving on to discuss the facilitators of effective programme implementation (Figure 15), the majority of headteachers indicated that the initial PALS-UK training for teachers (91.3%, 21 of 23), the four-week pupil training (73.9%, 17 of 23), the selection of reading books provided (65.2%, 15 of 23), the PALS-UK manual (56.5%, 13 of 23), and the half-day top-up training (52.2%, 12 of 23) were all useful to a great extent for the effective implementation of the programme. Furthermore, the vast majority of other headteachers in these categories indicated that these resources or activities were somewhat useful, rather than useful to a very little extent or not at all.

Figure 15: Headteachers' perceptions of facilitators to effective implementation of PALS-UK



	The initial PALS-UK training session for Year 5 teachers	The top-up training session for Year 5 teachers	The four weeks of training for Year 5 pupils	The PALS-UK manual	The ongoing support provided by the delivery team	The additional support provided by the delivery team	Video resources	Email resources	The selection of reading books provided	Other (please specify)
To a great extent	21	12	17	13	11	10	5	5	15	2
Somewhat	2	9	6	9	12	13	10	12	6	0
Very little	0	2	0	0	0	0	1	0	0	0
Not at all	0	0	0	0	0	0	1	0	0	1
Don't know	0	0	0	1	0	0	6	6	2	2

Headteacher endline survey question 25: n = 23 for items 1–9 (from left to right); n = 5 for item 10. 'Please indicate the extent to which you consider the following factors to have been useful in facilitating the effective implementation of the PALS-UK programme.'

[21] Given that these were open text responses in the survey, wider context to these responses cannot be provided by the evaluators.

Other factors were also considered facilitators by headteachers: all considered the ongoing support by the delivery team (100%, 23 of 23) and the additional support provided by the delivery team (100%, 23 of 23) to be at least somewhat useful for effective implementation. However, there was less agreement about the usefulness of video and email resources: six respondents in each of these categories indicated that they did not know (26.1%, 6 of 23 in each category) with the most common response being that they were somewhat useful (43.5%, 10 of 23 for video resources; 52.2%, 12 of 23 for email resources). Regarding video resources, one headteacher felt them to be only useful to a very little extent or not at all. However, it should be noted that video resources (as well as email correspondence) were directed only to teachers implementing PALS-UK. One case study interviewee highlighted the video resources in particular as an enabler to effective implementation, along with the manual:

> *'Definitely the video from the children and the pack we received with the manual [were enablers]: it literally had a script of things we could say to the children and help them with things they were finding hard' (TS1-1).*

Finally, of the few respondents that indicated 'other' with regards to facilitators, responses included the provision of reading books and resources on the memory stick.

Support for school staff in delivering PALS-UK

As described in the Intervention section, schools received two main forms of support: (1) ongoing support from the PALS-UK delivery team offered to all schools delivering the intervention and (2) additional support from the delivery team provided on an as-needed basis. This support mainly took the form of videos, emails, and just-in-time support from RAs. The perceived quality and usefulness of these forms of support are discussed and assessed in this section.

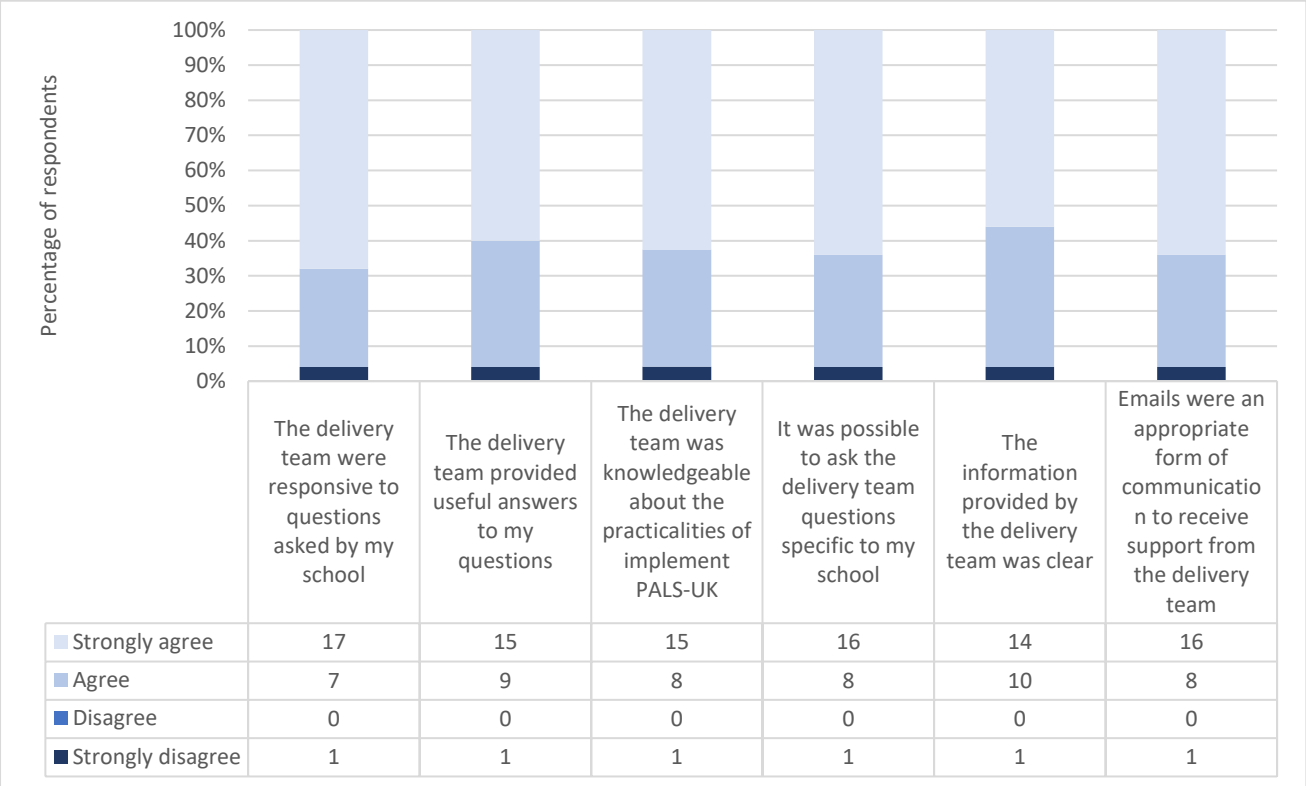
Perceptions of ongoing support from the delivery team

Teachers that had indicated that they made use of the ongoing support from the delivery team were then asked to report on the perceived quality and usefulness of this support (Figure 16). Overall, insights from the survey revealed that the majority of teachers perceived the ongoing support to be of high quality and useful to support and enhance the delivery. All but one teacher agreed or strongly agreed that the delivery team were responsive to questions, had provided useful answers to questions, were knowledgeable about the practicalities of implementing PALS-UK, that it was possible to ask school-specific questions, that the information provided was very clear, and that emails were an appropriate form of communication to receive from the delivery team. The one 'strongly disagree' response found across the items listed in Figure 16 came from the same school, indicating that it had not found the ongoing support useful. Nonetheless, the overarching message from the endline survey was that this ongoing support was useful for PALS-UK delivery. This was supported by insights from the case study interviews:

> *'We knew that if we had any questions, we could just email them' (TS3-1).*

> *'We had a lady who came to visit us … They've been amazing with keeping contact and answering our questions. They were really helpful' (TS1-1).*

Figure 16: Teachers' perceptions of the quality and usefulness of ongoing support provided by the PALS-UK delivery team



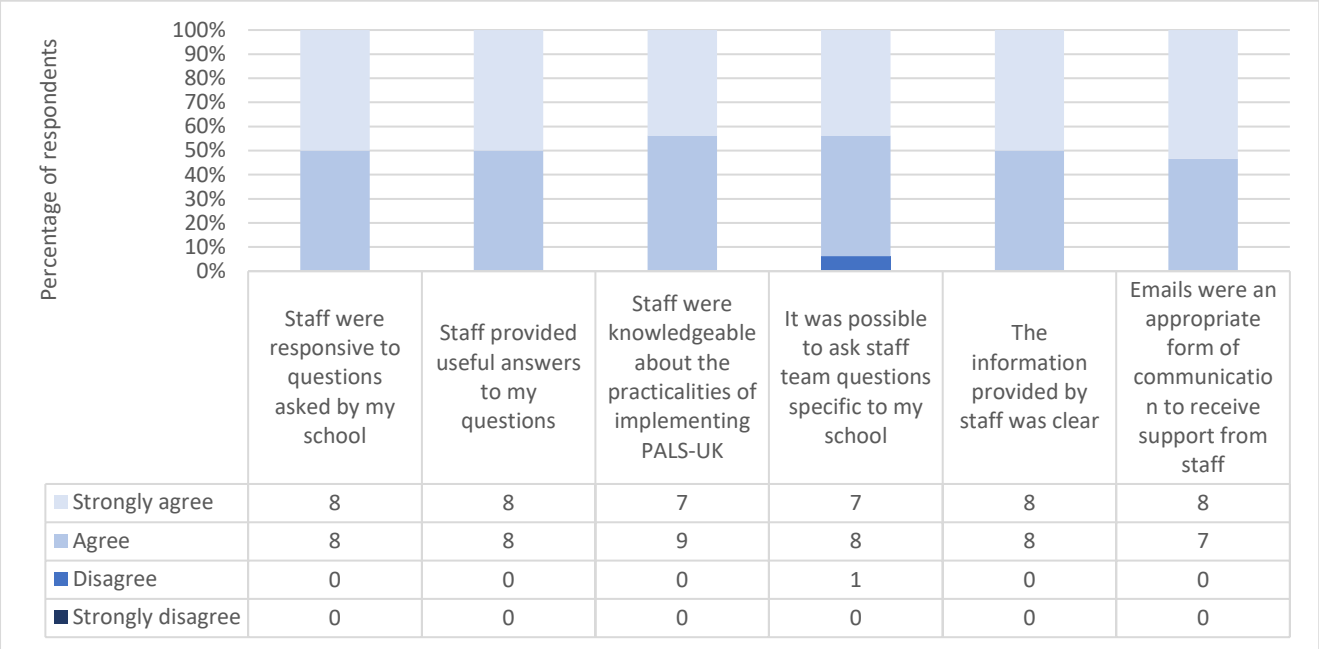
	The delivery team were responsive to questions asked by my school	The delivery team provided useful answers to my questions	The delivery team was knowledgeable about the practicalities of implement PALS-UK	It was possible to ask the delivery team questions specific to my school	The information provided by the delivery team was clear	Emails were an appropriate form of communicatio n to receive support from the delivery team
Strongly agree	17	15	15	16	14	16
Agree	7	9	8	8	10	8
Disagree	0	0	0	0	0	0
Strongly disagree	1	1	1	1	1	1

Teacher endline survey question 19: n = 25 for items 1–2 and 4–6 (from left to right); n = 24 for item 3. 'You have indicated that your school received ongoing support from the PALS-UK team to deliver PALS-UK during the trial. Please indicate the extent to which you agree or disagree with the following statements.'

Perceptions of additional support provided by the delivery team

The teachers that reported that they had used the additional support provided by the PALS-UK delivery team in the endline survey were then asked for their views on the usefulness of this additional support for delivery. As shown in Figure 17, views were positive: all teachers agreed or strongly agreed that staff were responsive to questions, provided useful answers, were knowledgeable about the practicalities of implementing PALS-UK, that the information provided was clear, and that emails were an appropriate form of communication. Additionally, the vast majority of teachers agreed that it was possible to ask school-specific questions (93.8%, 15 of 16), with just one teacher disagreeing with this statement.

Figure 17: Teachers' views on the usefulness of the additional support provided by the PALS-UK delivery team



	Staff were responsive to questions asked by my school	Staff provided useful answers to my questions	Staff were knowledgeable about the practicalities of implementing PALS-UK	It was possible to ask staff team questions specific to my school	The information provided by staff was clear	Emails were an appropriate form of communication to receive support from staff
Strongly agree	8	8	7	7	8	8
Agree	8	8	9	8	8	7
Disagree	0	0	0	1	0	0
Strongly disagree	0	0	0	0	0	0

Teacher endline survey question 22: n = 16 for items 1–5 (left to right); n = 15 for item 6. 'You have indicated that you have used the additional support available to deliver PALS-UK. We would like to know more about your experiences with other members of the delivery team i.e. staff who conducted the observations. Please indicate the extent to which you disagree or agree with the following statements.'

Other support for PALS-UK delivery

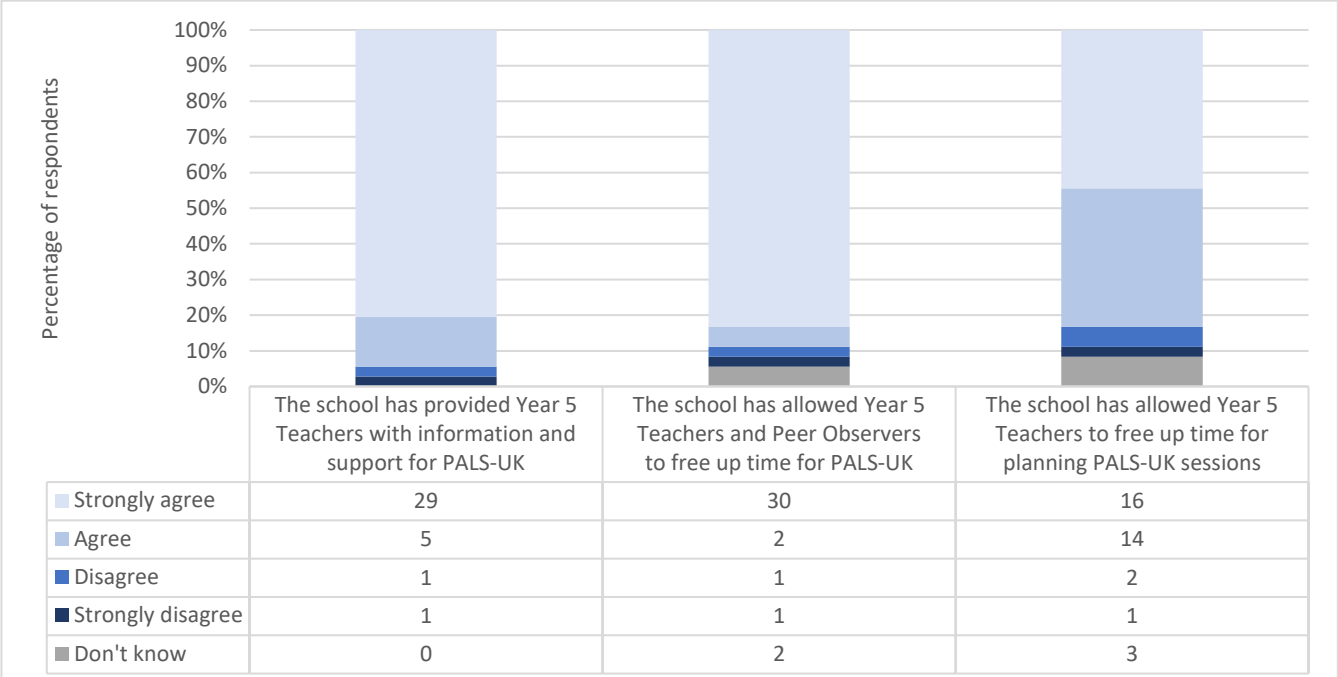
Teachers in intervention schools were asked whether they sought any support to implement PALS-UK from within their school or from other schools or colleagues. Just over 30% indicated that they had sought such support (30.8%, 12 of 39).[22] Thus, only a minority of surveyed teachers reported seeking support beyond the support provided by the PALS-UK delivery team.

As shown in Figure 18, headteachers in intervention schools indicated at baseline and endline that their school had provided delivery support to Year 5 teachers: the majority of headteachers agreed or strongly agreed that they had provided such support—94.4%, 34 of 36 at baseline; 95.7%, 22 of 23 at endline. Similarly, the majority of headteachers reported that they had allowed teachers and observers to free-up time for PALS-UK (88.9%, 32 of 36 at baseline; 95.7%, 22 of 23 at endline). While a lower proportion of headteachers strongly agreed with this, the majority agreed or strongly agreed that they had allowed Year 5 teachers to free-up time to plan PALS-UK sessions at baseline (83.3%, 30 of 36).

[22] Teacher endline survey question 23, n = 39: 'Did you seek any support to implement PALS-UK outside of the delivery team (i.e. within your school or with other schools/colleagues)?'

Figure 18: Headteachers views on support provided to teachers in the school to deliver PALS-UK at baseline and endline

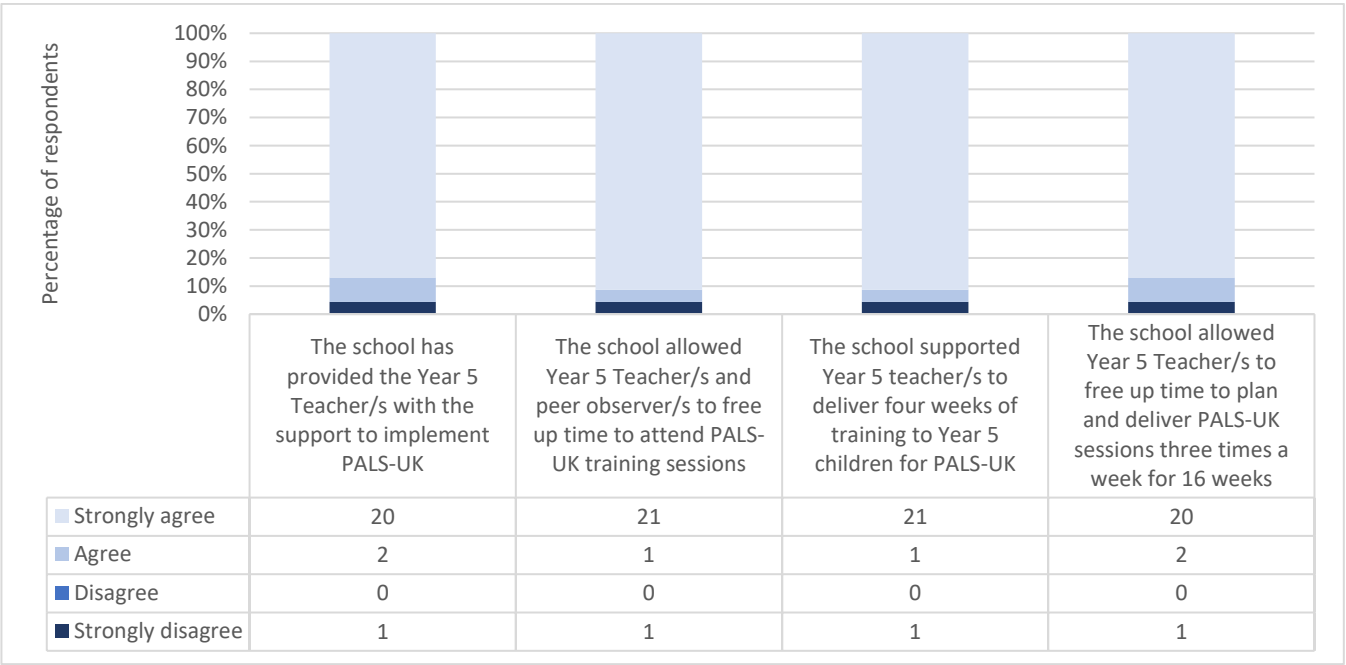
Figure 18a: Headteachers' views at baseline



	The school has provided Year 5 Teachers with information and support for PALS-UK	The school has allowed Year 5 Teachers and Peer Observers to free up time for PALS-UK	The school has allowed Year 5 Teachers to free up time for planning PALS-UK sessions
Strongly agree	29	30	16
Agree	5	2	14
Disagree	1	1	2
Strongly disagree	1	1	1
Don't know	0	2	3

Headteacher baseline survey question 28, n=36 across all three items. 'Please indicate how far you agree with the below statements. For each statement below, please select one of the options.'

Figure 18b: Headteachers' views at endline



	The school has provided the Year 5 Teacher/s with the support to implement PALS-UK	The school allowed Year 5 Teacher/s and peer observer/s to free up time to attend PALS-UK training sessions	The school supported Year 5 teacher/s to deliver four weeks of training to Year 5 children for PALS-UK	The school allowed Year 5 Teacher/s to free up time to plan and deliver PALS-UK sessions three times a week for 16 weeks
Strongly agree	20	21	21	20
Agree	2	1	1	2
Disagree	0	0	0	0
Strongly disagree	1	1	1	1

Headteacher delivery survey question 20, n = 23 across all four items. 'Please indicate the extent to which you disagree or agree with the following statements.'

At endline, the vast majority of headteachers agreed or strongly agreed that the school had supported teachers to deliver the four-week training to Year 5 pupils (95.7%, 22 of 23) and that they had allowed teachers to plan and deliver PALS-UK sessions three times a week for 16 weeks (95.7%, 22 of 23). Just one headteacher strongly disagreed with the

statements listed in Figure 18b indicating that issues related to school's supporting staff to deliver PALS-UK were rare. It should be noted that the same headteacher strongly disagreed with the four statements shown in Figure 18b.

Sentiments from the case studies also indicated that support from staff in the school for PALS-UK delivery was strong, corroborating insights from the surveys:

> *'Across the school everyone was really interested in it and we were able to do a staff meeting about it. And everyone commented on how engaged the children seemed' (TS3-1).*

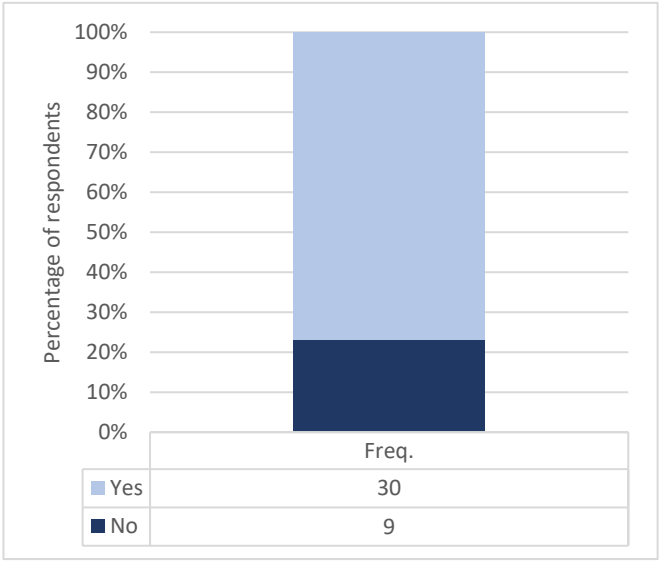
> *'They've been really good too. Our peer observer oversaw the implementation and wanted to trial it with other children at the school. They've really been on board with it' (TS1-1).*

Continuing with PALS-UK

At endline, teachers and headteachers in intervention schools were asked whether they would like to continue delivering PALS-UK beyond the conclusion of the trial: the majority of teachers indicated that they would like to if the programme was available in the future (77.0%, 30 of 39); the majority of headteachers indicated that they planned to continue delivering PALS-UK (82.6%, 19 of 23), but only a minority indicated that this would be on an unconditional basis (30.4%, 7 of 23): just over half of the headteachers indicated that they planned to continue delivering PALS-UK, but only if certain conditions were met (52.2%, 12 of 23).

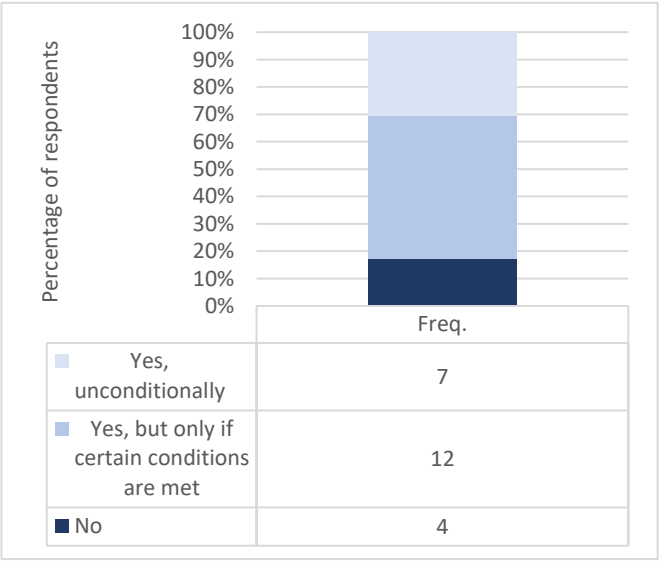
Figure 19: Teacher and headteacher views on whether they would like to continue delivering PALS-UK beyond the end of the trial

Figure 19a: Teachers



Teacher endline survey question 29, n = 39: 'If the programme was available to your school in the future, would you like to continue implementing PALS-UK?'

Figure 19b: Headteachers



Headteacher endline survey question 29, n = 23: 'Does your school plan to continue delivering PALS-UK for Year 5 pupils after the trial finishes, once your school fully re-opens?'

Among the headteachers that indicated that they planned to continue with PALS-UK if certain conditions were met, conditions mentioned included the school's perceived need for PALS-UK (n = 7), the eventual financial costs of the programme (n = 5), and whether or not other reading programmes were running in the school (n = 4). One other headteacher mentioned in their survey response that their school's continuation with PALS-UK is dependent on whether they can train up other staff members in the approach, given that the teacher delivering PALS-UK in their school was moving to another position. Insights from the case studies here suggested that schools wanted to make adaptations to the programme before continuing with implementation, for example, the number of times delivered per week:

> *'Yes [we will likely continue implementing PALS-UK] but probably having it twice per week rather than three times. Or possibly breaking down the sections so it's not so repetitive for them' (TS3-1).*

Reasons for not continuing mentioned by headteachers in the endline survey included the programme not being beneficial enough to Year 5 pupils to justify continuing (n=2), the school lacking the resources (n = 1), and delivery of other reading interventions in the school (n = 1). Other reasons included teachers trained in PALS-UK no longer being available to do so (n = 1) and elements of the programme being adapted to create a bespoke reading programme to address other facets of reading, for example, inference and word meaning (n = 1).

Perceived outcomes associated with PALS

Box 3: Perceived outcomes—key findings

Responses to staff surveys and data from the in-depth interviews in case study schools around the perceived impact of PALS-UK show the following:

1. Responses provided in the post-delivery surveys show that Year 5 teachers' perceptions on the impact on pupils in terms of improved reading attainment is mixed.

2. Interview responses show that school staff were more positive about the impact on reading confidence and reading fluency than on the impact on reading comprehension, particularly immediately after the 20-week implementation had ended.

3. Schools felt that PALS-UK had positive impacts on a wide range of outcomes that are not captured by the logic model, including improvements made in teachers' reading instructions, pupils' vocabulary, and improvement in independent pupil learning.

4. Survey and interview responses suggest that closure of schools and the movement to online learning as a result of national lockdown from March to September 2020 had a negative impact on pupils' reading skills. There was also evidence suggesting that any gains in reading linked to the use of PALS-UK were lost through to end of the trial due to the COVID-19 pandemic.

While the Impact evaluation section has discussed pupil-level outcomes based on reading attainment measures, the following section looks at the perceived impact of PALS-UK using data collected as part of the IPE. School staff in treatment schools were asked about the impact of PALS-UK on the development of pupils shortly after the programme ended in the post-delivery survey. This was to explore intended and unintended programme consequences. In the follow-up IPE round, we surveyed Year 6 teachers to assess the impact of COVID-19 on pupil's reading skills in the next academic year and to assess to what extent programme effects on pupil's is observed in the next academic year. In addition to the surveys, the IPE also captured information about the perceived impact of PALS-UK on pupils through two rounds of interviews with staff from the case study schools. The findings on perceived programme impact are discussed in turn below.

Perceived outcomes of using PALS-UK

An exploration of the perceived impact of PALS-UK was conducted using responses in the post-delivery survey and in the follow-up survey as well as using responses via interviews with case study schools. These are discussed for the predicted outcomes in the logic model (Figure 1).

Teachers in intervention schools were asked several questions in the surveys about the impact of PALS-UK on pupils. At the end of the 2019/2020 school year, two months after programme delivery was completed, we asked Year 5 teachers how they perceived PALS-UK had impacted pupils. Around half of the teachers (53.8%, 21 of 39) *observed improvements in the reading ability of Year 5 pupils.* However, 17.95% (7 of 39) reported that they 'disagree', while 28.21% (11 of 39) 'strongly disagree' that PALS-UK would lead to improvement in reading attainment. This is suggesting that the evidence from the surveys on the observed programme impact on pupils is mixed.

It is worth noting that while the post-delivery surveys undertaken in June to August 2020 and November 2020 to March 2021 did not provide information about the impact of PALS-UK on oral reading fluency, confidence, or reading comprehension, in the post-delivery interviews undertaken between May and June 2020 we asked school staff how each of these were affected in the period immediately after delivery ended. The PALS-UK logic model specifies that the programme's route to impact is linked, first, to improving oral reading fluency and reading comprehension (see Figure 1), which will in turn lead to improvement in pupil's reading attainment.

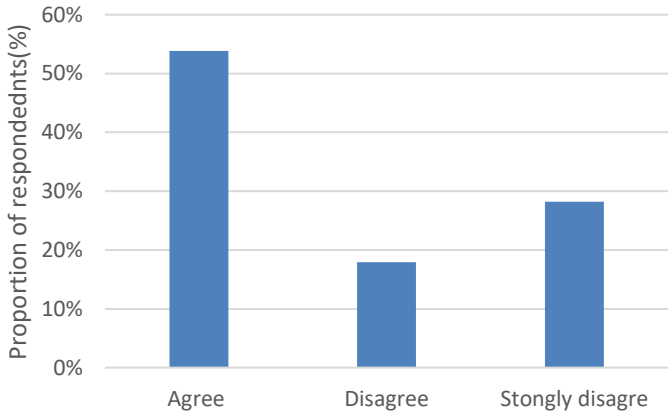
Although it is worth noting that only a small number of interviews were conducted in case study schools at endline—and such schools, by participating, are likely to be more engaged and reflect positively on PALS-UK—some interesting impacts of the PALS-UK programme were highlighted. Overall, evidence from the first round of interviews suggests that participation in the programme had a perceived positive impact on *reading fluency* and *confidence in reading*, however, evidence for an impact on reading comprehension was mixed.

The vast majority of school staff in the first round of interviews suggested that they observed biggest improvements in reading fluency (HTS2-1, T1S2-1, HTS3-1). One headteacher observed that 'there was massive progress in every child' and that the progress was mainly observed in their reading fluency scores (TS1-1).

Similarly, data from the in-depth case study interviews indicated that PALS-UK was perceived to have helped to increase pupils' confidence in reading (in that pupils were less shy to read in front of others: TS1-1; HTS2-1) or they felt PALS-UK activities had increased pupils' engagement with reading activities (HTS2-1). For example, the following quote summarises the views of staff from case study school two on the impact they observed PALS-UK to have on pupils' confidence:

> *'The fact that as a less able child you can challenge and ask a more able child questions empowers them that they can do it and makes them know that they're just as good a reader as they are' (HTS2-2).*

Figure 20*:* Teachers' views on the PALS impact on reading attainment



Treatment teacher endline post-delivery survey, n = 39. 'Please indicate the extent to which you disagree or agree with the following: have observed an improvement in the reading ability of Year 5 pupils as a result of implementing PALS-UK, over and above what I would normally expect.'

Teachers and headteachers from case study schools also commented that they observed improvement in reading comprehension, but views were not uniformly held. One headteacher expressed that they felt pupils did not fully improve comprehension (HTS-1), and another teacher mentioned that pupils required 'a lot of guidance' with comprehension unlike with reading fluency (TS3-1).

Unintended consequences on the perceived impact of PALS-UK linked to COVID-19

In the following academic year, specifically in December 2020 (around the same time at which PALS-UK pupils were tested), we asked headteachers about the impact of COVID-19 on Year 6 pupils' literacy development. The majority reported observing negative (76.5%, 13 of 17) or extremely negative (11.8%, 2 of 17) impacts on Year 6 pupil's reading fluency and comprehension due to the movement to remote learning as part of lockdown and the social distancing measures. The COVID-19 outbreak was also felt to have had a negative impact on pupils' confidence in reading, with majority of teachers reporting negative (70.6%, 12 of 17) or extremely negative impact (5.9%, 1 of 17). When control schools were asked the same two questions, 62.5% (5 of 8) and 75% (6 of 8) of headteachers reported that COVID-19 restrictions negatively affected pupils' reading fluency and confidence, respectively. The fact that the vast majority of headteachers reported seeing negative impact in pupils' reading skills and confidence may serve as an indication the pandemic may have outweighed any positive programme impact from the previous academic year. It should be noted that it is not possible to compare perceived changes in pupils' reading outcomes between treatment and control schools in the period between the end of the PALS-UK implementation and the COVID-driven IPE phase as the IPE did not

collect data on the perceived changes in pupils' reading outcomes in control schools at either of these time points as this was out of the scope of the evaluation.

These survey findings were supported by the analysis of the second round of interviews, which indicated that the COVID-19 outbreak negatively impacted pupils reading fluency and comprehension. The majority of respondents felt that any early improvement in reading associated with participation in PALS-UK may have dropped during the COVID-19 lockdowns (HTS2-2; HTS4-2; TS6-2; TS7-2). The negative change they felt could be attributed to children being less likely to read when not in school or when not surrounded by other children (HTS4-2), or because of not having an opportunity to read in a pair (TS6-2). One respondent did not feel that reading fluency was as negatively affected as reading comprehension (HTS4-2).

Additional unintended consequences of using PALS-UK

Improved teacher knowledge on reading instruction

Surveyed and interviewed teachers and headteachers identified that the programme had a positive impact on Year 5 teachers in terms of improved knowledge on reading instruction. In the endline survey, 87.08% (34 of 39) of teachers 'agreed' or 'strongly agreed' that as a result of PALS-UK they have observed improvement in how they teach reading. Some interviewees also reported that they observed positive impacts of PALS-UK on the knowledge and skills of Year 5 teachers (TS3-1).

There were also some additional unintended impacts of the programme not specified in the hypotheses or the logic model. For example, a couple of teachers explained how the intervention was implemented not only in the Year 5 classroom but across different year groups at the school.

> *'The positive was that it was a new initiative and action-based research is a revitalising opportunity to the whole school. It gave a new a focus to that particular year group' (HTS3-1).*

A new sense of independence in terms of pupil learning

One interviewee attributed a new sense of independence in terms of pupil's learning to their participation in the programme. The correction procedure was identified by one interviewee as encouraging independent learning since it 'enabled the children to identify their own mistakes' (T1S2-1). One teacher noted that the development of new and important skills such as peer feedback, facilitated independent learning (TS3-1). One case study school indicated that PALS-UK was observed to have helped increase *vocabulary*:

> *'We started to develop some work around vocabulary and how you capture certain words. So teachers began writing words on the board that they didn't really understand. We also included Year 6 children in a booster group and it was noticeable that their vocabulary wasn't as developed as some of the children in the PALS programme' (HTS2-1).*

While the Perceived Impact section has discussed perceived pupil-level outcomes—including what areas of language showed the most improvement—we did not explore the varying perceived impact for different groups of pupils. This was out of the scope of this evaluation.

Usual practice

Box 4: Usual practice key findings

Evidence gathered from surveys and interviews with headteachers and teachers across the intervention and control arms at baseline, endline, and during the COVID-19-driven phase show that, with regards to usual practice:

1. At baseline, the majority of schools had dedicated guided reading time for Year 5 pupils, which tended to be last between one and three hours per week.
2. Just under half of headteachers across intervention and control schools indicated that they had reading programmes available in their school at baseline; the most commonly mentioned being Accelerated Reader.
3. At endline, the majority of headteachers in control schools reported that they had dedicated guided reading time for Year 5 pupils, often lasting between one and two hours, or more, and implemented in whole class format.
4. At endline, only a minority of headteachers in control schools reported that reading programmes were available in their school, the most commonly mentioned again being Accelerated Reader.
5. Only a small number of intervention and control schools were engaged in any specific programme targeted at improving the reading skills of Year 6 pupils.
6. Other programmes and support besides PALS-UK were being delivered in intervention schools to the Year 6 cohort but the level and frequency of this is unclear.
7. While the majority of headteachers in interventions schools confirmed that the school provided additional support or resources for improving reading comprehension and fluency to Year 6 pupils, the evidence was mixed in control schools as only half of the headteachers confirmed that they provided additional support or resources and the survey response rate in the first place was lower in control schools, meaning that the results should be interpreted with caution.

For the purposes of this evaluation, usual practice was explored in multiple ways: looking at the context of what schools were doing prior to delivering PALS-UK and also at what schools were doing instead of delivering PALS-UK (in control schools) or alongside PALS-UK (in intervention schools). To contextualise findings further, we also explored what schools were doing since they stopped delivering and before testing took place. All of these allow a better understanding of context. These will be discussed in turn below.

Usual practice in all schools at baseline

Exploring usual practice in all schools at baseline will focus on two main elements: (1) whether schools had dedicated guided reading time at baseline and (2) whether schools had taken part in other reading-focused programmes at baseline.

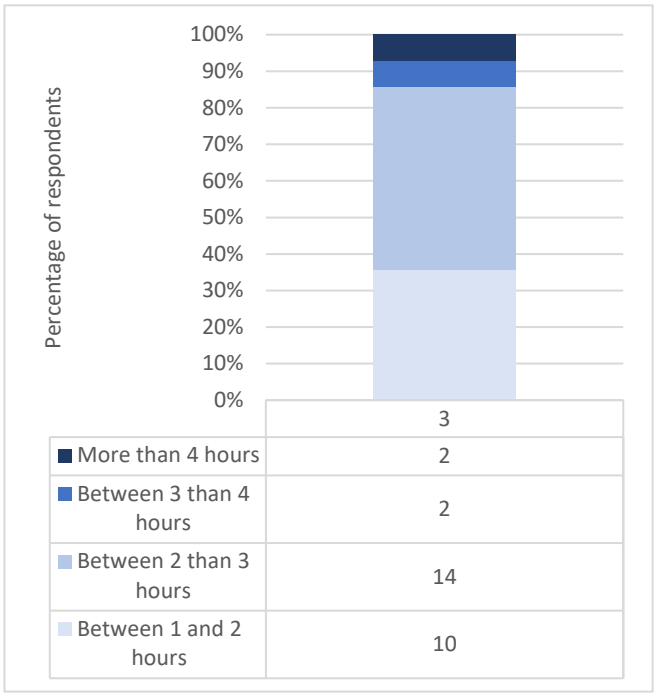
Guided reading practices at baseline

Looking at whether schools had dedicated guided reading time, responses to the baseline headteacher surveys across intervention and control schools indicated that the majority of schools across both conditions did have this time built into the timetable (86.1%, 31 of 36 intervention headteachers; 83.8%, 31 of 37 control headteachers).

As shown in Figure 21, of those headteachers that indicated that there was dedicated guided reading time, the majority across intervention and control schools indicated that this time amounted to between one and three hours per week (77.4%, 24 of 31; 80.6%, 25 of 31 respectively). Only a minority said it was three hours or more per week (12.9%, 4 of 31 in intervention schools; 6.5%, 2 of 31 in control schools).

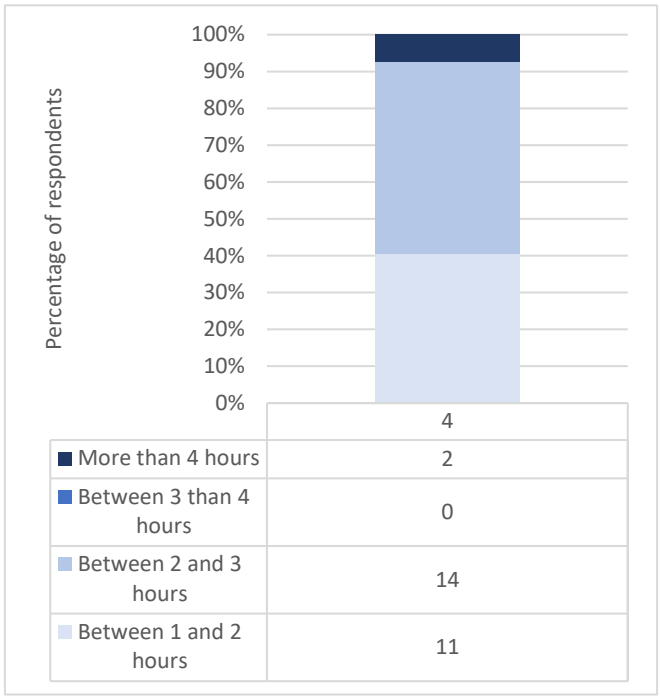
Figure 21: Usual practice related to guided reading time in intervention and control schools at baseline

Figure 21a Intervention schools



Headteacher intervention baseline survey question 5, n = 31. 'How much guided reading time per week is there in Year 5 pupils' timetable?'

Figure 21b Control schools



Headteacher control baseline survey question 5, n = 31. 'How much guided reading time per week is there in Year 5 pupils' timetable?'

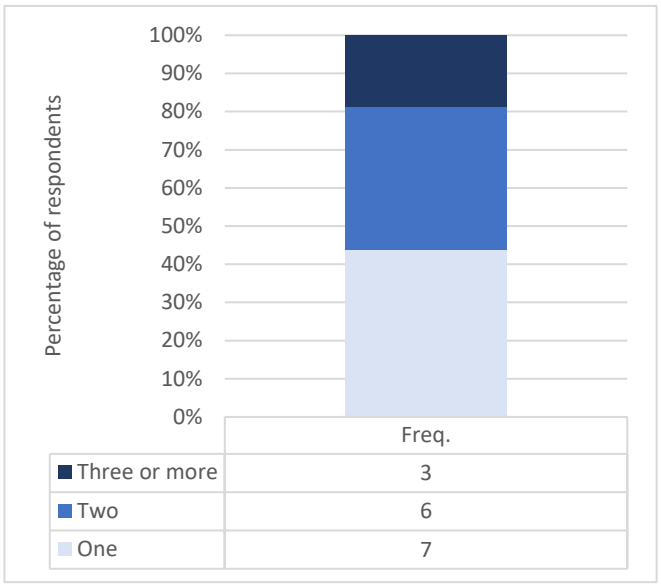
Overall, then, usual practice related to dedicated guided reading time was somewhat similar across intervention and control schools at baseline.

Other reading-focused programmes in use at baseline

Looking next to other reading-focused programmes in schools at baseline (Figure 22), a similar proportion of headteachers across intervention and control schools indicated in the surveys that other reading programmes apart from PALS-UK were available in their school (44.4%, 16 of 36 in intervention schools; 40.5%, 15 of 37 in control schools). Of these, the majority indicated that one or two reading programmes were available (intervention: 81.3%, 13 of 16; control: 86.7%, 13 of 15). An additional three headteachers in intervention schools (18.8%) and two in control schools (13.3%) indicated that three or more programmes were available.

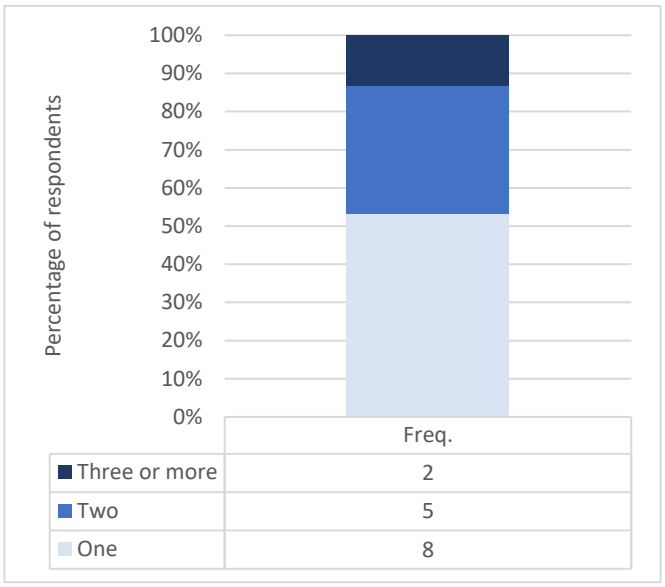
Figure 22: Headteachers' response on the number of other reading programmes available in the school at baseline across intervention and control schools

Figure 22a Intervention



Headteacher intervention baseline survey question 7, n = 16. 'How many reading programmes are there currently available for Year 5 pupils in your school?'

Figure 22b Control



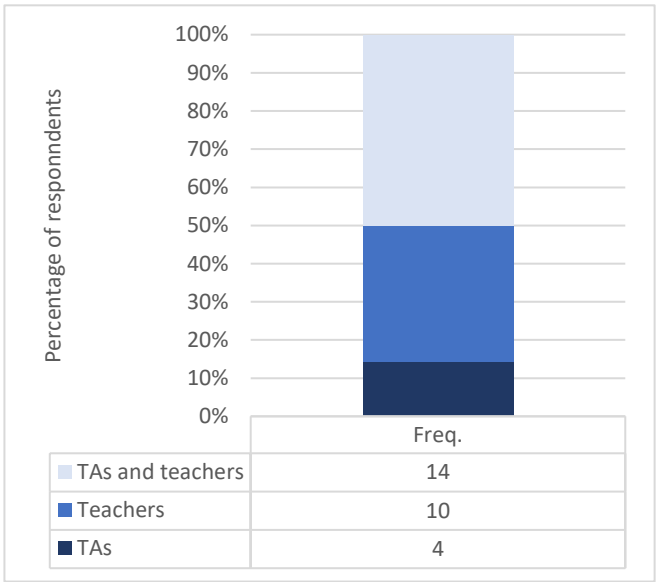
Headteacher control baseline survey question 7, n = 15. 'How many reading programmes are there currently available for Year 5 pupils in your school?'

Among the programmes mentioned, the most common was Accelerated Reader (n = 7 times mentioned in intervention schools; n = 9 times mentioned in control schools). Other programmes mentioned less frequently included Catch Up Reading, Reading Plus, Boosting Reading at Primary, Resilient Reader, Lexia, Toe by Toe, Book Talk, and Reading Rainbow. One intervention headteacher mentioned that PALS-UK was available in their school, although it is possible that they meant the programme was available through the trial.

When asked about who delivers these programmes (Figure 23), the majority said it was a combination of teachers and TAs (n = 14 in intervention schools; n = 11 in control schools) or teachers only (n = 10 in intervention schools; n = 7 in control schools). Across both arms, it was least common for headteachers to report that the programmes were delivered solely by TAs (n = 4 across both arms).

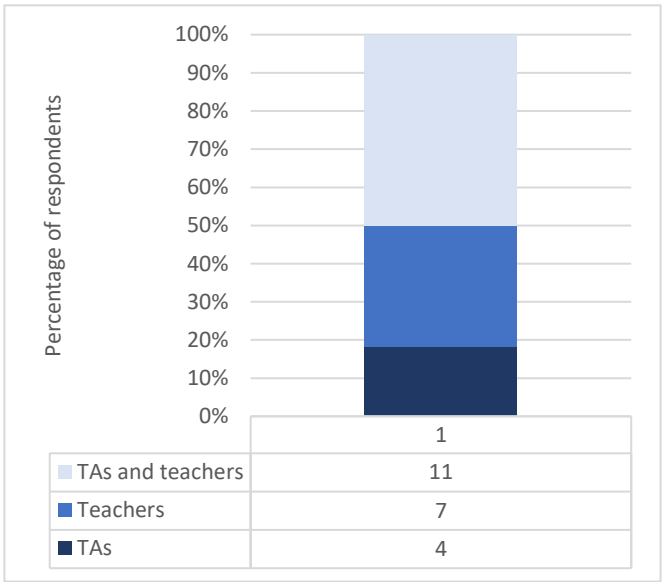
Figure 23: Headteacher response on who delivers reading programmes in intervention and control schools at baseline

Figure 23a Intervention



Headteacher intervention baseline survey questions 9, 15, and 21, n = 16. Responses do not add up to 16 as this question was repeated up to three times depending on the number of programmes available in the school. Question: 'Who normally delivers this programme?'

Figure 23b Control



Headteacher control baseline survey questions 9, 15 and 21, n = 15. Responses do not add up to 15 as this question was repeated up to three times depending on the number of programmes available in the school. Question: 'Who normally delivers this programme?'

When asked how the programme is delivered, responses were also mostly similar across intervention and control schools. Most said via small group instruction (n = 8 in intervention schools; n = 7 in control schools) and to a slightly lesser extent via one to one instruction (n = 3 in intervention schools; n = 4 in control schools). However, it was more common for headteachers in intervention schools to have programmes delivered via whole school instruction than in control schools (n = 11 and n = 5 respectively).[23]

In terms of how long these programmes lasted, most reported more than 20 weeks (n = 21 in intervention schools; n = 16 in control schools). A smaller number of headteachers indicated that the programmes lasted five to ten weeks (n = 6 in intervention schools; n = 4 in control schools) and to an even lesser extent 11 to 20 weeks (n = 2 in intervention schools; n = 3 in control schools). One headteacher in a control school reported that they do not really implement the programme activities.[24]

When asked about how frequently programme activities take place, most said once a day (n = 13 in intervention schools; n = 11 in control schools) or at least once per week (n = 15 in intervention schools; n = 10 in control schools). Just one headteacher apiece in intervention and control schools indicated that the programmes were delivered in every literacy lesson.[25]

When asked to summarise the purpose of these reading programmes, open text responses to the baseline survey included: to teach reading and comprehension skills (n = 8), to improve reading fluency (n = 2), to encourage regular independent reading (n = 2), to improve engagement with reading (n = 6), to increase children's stamina for reading, to

[23] Headteacher intervention baseline survey questions 10, 16 ,and 22 (n = 16). Headteacher control baseline survey questions 10, 16, and 22 (n = 15). Responses do not add up to these figures as this question was repeated up to three times depending on the number of programmes available in the school. Question: 'How is the programme delivered?'
[24] Headteacher intervention baseline survey questions 11, 17, and 21 (n = 16). Headteacher control baseline survey questions 11, 17, and 21 (n = 15). Responses do not add up to these figures as this question was repeated up to three times depending on the number of programmes available in the school. Question: 'How long does the programme last?'.
[25] Headteacher intervention baseline survey questions 12, 18, and 22 (n = 16). Headteacher control baseline survey questions 12, 18, and 22 (n = 15). Responses do not add up to these figures as this question was repeated up to three times depending on the number of programmes available in the school. Question: 'How frequently do programme activities take place?'

develop a reading culture in the school, to promote accelerated reading learning (n = 2), and for reading and phonics catch up (n = 2).

Overall, a range of programmes focused on reading were available in intervention and control schools at baseline. The most commonly mentioned programme available across intervention and control schools at baseline was Accelerated Reader. Most were delivered by a combination of TAs and teachers (or teachers alone), delivered in small groups, more than 20 weeks in length, and delivered at least once per week or even daily.

Usual practice in control schools during the trial

Control schools during the trial were asked a number of questions to understand their usual practices. Specifically, they were asked about usual practices related to reading learning in their school as well as the availability of reading-focused programmes in their school.

Looking first at practices related to reading learning, when asked at endline whether the school had dedicated guided reading time in the timetable for Year 5 pupils, as at baseline the vast majority of headteachers in control schools indicated that this was the case (90.9%, 20 of 22).[26] When asked about the amount of guided reading time there was per week for Year 5 pupils, headteachers commonly indicated that this was between one and two hours or more than two hours (45.0%, 9 of 20 headteachers); only two headteachers indicated that it was less than one hour per week (10.0%, 2 of 20).[27] Headteachers were also asked about how this guided reading time was implemented. A range of methods were mentioned as open text responses, including whole-class guided reading and discussion (n = 7), carousel activities (n = 4), group reading activities (n = 5), paired reading (n = 1), and independent reading (n = 1).

Both of the headteachers that indicated that the school did not have dedicated guided reading time reported that the school implements a supplementary reading practice or intervention, in these cases Accelerated Reader.[28]

Moving on to look at the availability of reading programmes in control schools during the trial, only a minority of headteachers indicated at endline that a reading programme was available in their school (31.8%, 7 of 22).[29] Of these, three indicated that one or two programmes were available while one indicated that there were three or more.[30] The programmes names were Accelerated Reader (n = 6), Bug Club Guided Reading (n = 1), Century AI (n = 1), CLIMB (n = 1), and Reciprocal Reading (n = 1).[31] Thus, as was the case at baseline, Accelerated Reader was the most commonly mentioned. Headteachers indicated that the purpose of these reading programmes was to foster a love of reading (n = 4), to improve reading comprehension (n = 3), and identify and address gaps in learning (n = 1).[32]

Most reported that the programmes were delivered by a combination of TAs and teachers (n = 7); some were delivered by teachers only (n = 3).[33] When asked how the programme is delivered, it was most common for headteachers to report either via whole group instruction (n = 3) or a combination of methods (for example, whole group with small group and independent sessions) (n = 3). Additionally, one headteacher indicated that the method of delivery was small group

[26] Headteacher control endline survey question 3, n = 22. Question: 'Does your school normally have dedicated guided reading time in the timetable for pupils in Year 5?'
[27] Headteacher control endline survey question 5, n = 20. Question: 'How much guided reading time per week is there normally in Year 5 pupils' timetable?'
[28] Headteacher control endline survey question 4, n = 2. Question: 'You have indicated that you do not use guided reading in your school. Do you implement any supplementary reading practice intervention? (If yes, please specify)'
[29] Headteacher control endline survey question 7, n = 22. Question: 'Has your school paid for or taken part in any specific programmes to improve the reading skills of Year 5 pupils during this school year (2019/2020)?'
[30] Headteacher control survey endline question 8, n = 7. Question: 'How many reading programmes are there currently available for Year 5 pupils in your school?'
[31] Headteacher control endline survey questions 9, 15 and 21 (n = 6). Responses do not add up to these figures as this question was repeated up to three times depending on the number of programmes available in the school. Question: 'What is the name of the reading programme? (Please fill in N/A if the programme does not have a name)'
[32] Headteacher control endline survey questions 10, 16 and 22 (n = 6). Responses do not add up to these figures as this question was repeated up to three times depending on the number of programmes available in the school. Question: ' What would you say is the main purpose of the reading programme?'
[33] Headteacher control endline survey questions 11, 17 and 23 (n = 6). Responses do not add up to these figures as this question was repeated up to three times depending on the number of programmes available in the school. Question: 'Who normally delivers this programme?'

instruction, one to one instruction, or independent work.[34] Interestingly, across all programmes mentioned by headteachers, the amount of time that the programme was delivered for was more than 20 weeks (n = 10).[35] Finally, when asked how frequently programme implementation took place, the majority of headteachers indicated that this was once a day (n = 7), with the others indicating that implementation took place at least once per week (n = 3).[36]

Overall, then, the reading programmes available in control schools as reported at endline tended to be delivered by a combination of TAs and teachers, delivered via who group instruction or a combination of techniques, 20 weeks or more in length, and delivered on a daily basis. Thus, the nature of these reading programmes on average tended to be quite different to the PALS-UK programme.

Usual practice after the return to school in September 2020 and before outcome testing

Control and intervention teachers were surveyed after the return to school in December and were asked about the use of reading attainment programmes (excluding PALS-UK) in their school between September 2020 and December 2020.

The evidence from the headteacher survey suggests that only a small number of schools were engaged in a specific structured programme or activity targeted at improving the reading skills of Year 6 pupils during this time, however, when we asked headteachers and teachers if the school offered any additional reading support to Year 6 pupils, the majority of them confirmed that they had (see Box 5). Box 5: Other reading programmes mentioned by control and intervention headteachers being delivered between September 2020 and December 2020 and Box 6 provide details on the specific programmes and support that were mentioned being delivered in the treatment schools. As can be seen below, a number of approaches were used: additional lessons focused around reading, one to one support, access to online resources, and additional TA time with targeted pupils. Overall, then, the evidence from the teacher and headteacher survey is suggesting that other programmes and support besides PALS-UK were being delivered in intervention schools to the Year 6 cohort but it is unclear how frequently.

To investigate this further, we interviewed school staff in case study schools about what reading support has been offered to Year 6 pupils. IPE data from the interviews with case study schools is suggesting that intervention schools were active in the new academic year following the lockdown. Interviewees noted that their school used other reading interventions besides PALS-UK with the Year 6 cohort following the return to school in September. Schools mentioned using Accelerated Reading (HTS2-2, TS5-2), unnamed small group interventions (TS5-2), Reading Explorers (TS5-2), and Racing to English (TS7-2).

We do not have systematic data across all schools, but by comparing intervention schools at baseline and at the follow-up surveys, it seems reasonable to infer that they were more active before they implemented PALS-UK (see Usual Practice in All Schools at Baseline), but less active than in the time PALS-UK was being delivered (see Usual practice in control schools during the trial) compared to the follow up period, that is, in the new academic year between September and December 2020.

Headteachers across the control schools were asked to indicate the activities the school had undertaken between September 2020 and December 2020. Only one headteacher confirmed that the school had taken part in a reading programme (12.5%, 1 of 8) (see Box 5). Considering the overall low response rate on the headteacher survey, this evidence should not be taken as an indicator that control schools were not active in the use of additional reading support to Year 6 pupils. When asked to describe any support or resources provided to support Year 6 reading development, half of the headteachers in control schools (4 of 8) confirmed that the school made available additional support (see Box 5).

[34] Headteacher control endline survey questions 12, 18, and 24 (n = 6). Responses do not add up to these figures as this question was repeated up to three times depending on the number of programmes available in the school. Question: 'How is the programme delivered?'

[35] Headteacher control endline survey questions 13, 19, and 25 (n = 6). Responses do not add up to these figures as this question was repeated up to three times depending on the number of programmes available in the school. Question: 'How long does the programme last?'

[36] Headteacher control endline survey questions 14, 20, and 26 (n = 6). Responses do not add up to these figures as this question was repeated up to three times depending on the number of programmes available in the school. Question: 'How frequently are the programme activities implemented in your school?'

Box 5: Other reading programmes mentioned by control and intervention headteachers being delivered between September 2020 and December 2020

Intervention schools

IPE data from the surveys with headteachers shows that 5 of 23 intervention headteachers (21.79%) reported using one (3 of 5) or two (2 of 5) reading programmes with Year 6 pupils—Reading Plus, Fiction Express, Oxford Owl, or First News Education. The only two respondents to indicate why they had implemented reading programmes said they were to improve **reading fluency** or to improve **reading comprehension**. Both of these respondents said that TAs and teachers are responsible for the delivery of these programmes. In terms of programme delivery, one respondent referenced whole group instruction, while another mentioned independent work at school. The two respondents also indicated that the reading programme lasted between five and ten weeks and more than 20 weeks.

Control schools

In the survey, only one of eight control headteachers (12.5%) reported their school had paid for or taken part in a programme to improve the reading skills of Year 6 pupils. The headteacher did not provide more information about the programme being implemented.

When asked whether the school provided any additional support or resources for improving reading comprehension and fluency to Year 6 pupils, 50% (4) said yes and 50% (4) said no.

Box 6: Additional support and resources offered in the schools for improving reading comprehension and fluency after the return to schools in September 2020 mentioned by headteachers and teachers from treatment schools

Intervention schools

The majority of headteachers in interventions schools confirmed that the school provided additional support or resources for improving reading comprehension and fluency to Year 6 pupils (88.2%, 15 of 17). When asked to provide more information on the form of these resources, headteachers mentioned:

- small group interventions;
- additional TA time for targeted pupils to support both reading fluency and comprehension;
- providing access to an online library to all children so they can continue reading books at home as books were not being sent home with children;
- additional lessons focused on reading and one to one reading support;
- additional booster sessions in the summer holidays offered to vulnerable learners; and
- extra reading sessions including one to one reading with children.

Similarly, most teachers (85.7%, 18 of 21) confirmed that the school offered additional support. This included:

- targeted daily reading with an adult;
- regular guided reading sessions with lots of written comprehension;
- reading books at home;
- more time being built into the school day;
- increased opportunities to read;
- teacher-planned activities around reading;
- more books being purchases;
- more online support being offered; and
- subscription to reading interventions.

When asked to what extent Year 6 pupils had engaged with such additional literacy support, 44.4% (8 of 18) said ‘somewhat’, 55.6% (10 of 18) said ‘to a great extent’.

Box 7: Additional support and resources offered in the schools for improving reading comprehension and fluency after the return to schools in September 2020 mentioned by headteachers from control schools

Control schools

When asked whether the school had provided any additional support or resources for improving reading comprehension and fluency to Year 6 pupils, half of the headteachers (50%, 4 of 8) confirmed they had. Of these, three confirmed that:

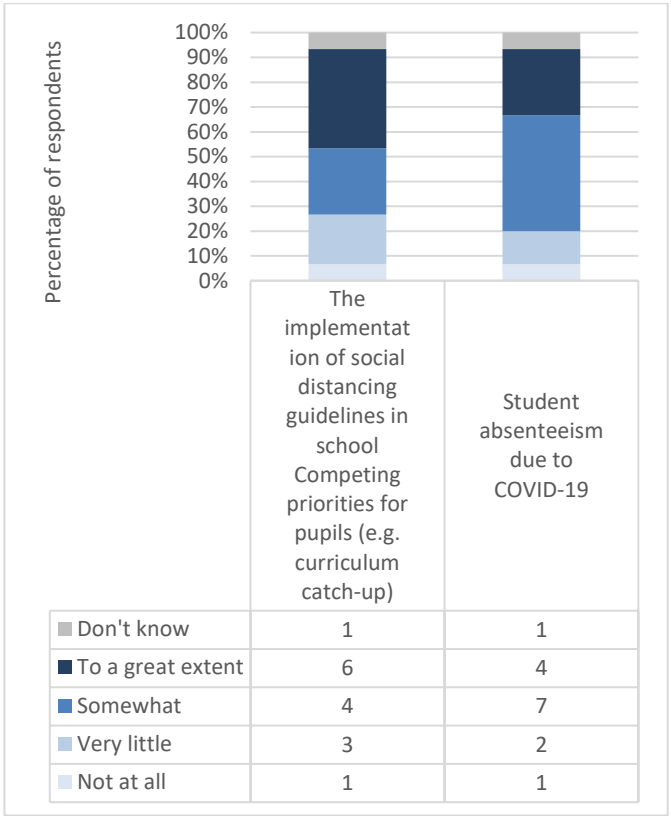
- pupils were taking part in a dedicated 12-week programme of small-group catch-up sessions and pupils had an individualised reading programme put in place via trained TAs;
- the schools offered daily pre-teaching sessions for children who have slipped below age related expectations; and
- the school made available increased one to one support with an adult, increased opportunities to read and discuss inference questions in class, and enhanced opportunities to link other areas of the curriculum to high quality texts.

Engagement with the additional support

School staff were asked to report on the extent they felt that pupils were engaging with the additional support. Survey data suggests that headteachers from the intervention schools felt that this was ‘high’ (53.3%, 8 of 15) or ‘somewhat high’ (46.7%, 7 of 15). Similarly, most headteachers in control schools that responded to question (75%, 3 of 4) felt that Year 6 pupils were engaging with the additional support. When asked about the reason why pupils were not engaging with the resources, the implementation of social distancing guidelines appeared to be slightly more important in intervention schools whilst student absenteeism appeared more important in control schools (see Figure 24).

Figure 24: Headteachers' responses on the reason Year 5 pupils were not engaging with additional support in intervention and control schools

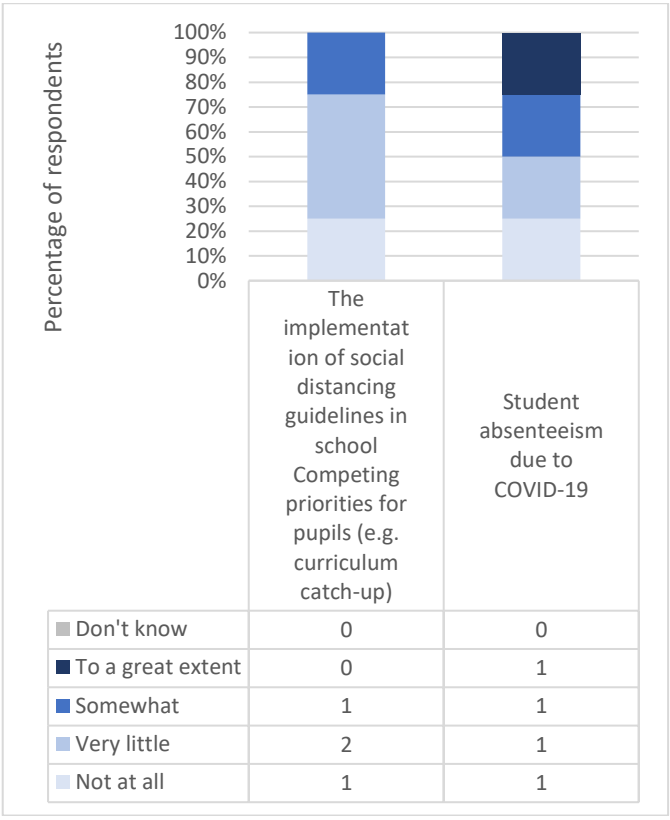
Figure 24a: Intervention



	The implementation of social distancing guidelines in school Competing priorities for pupils (e.g. curriculum catch-up)	Student absenteeism due to COVID-19
Don't know	1	1
To a great extent	6	4
Somewhat	4	7
Very little	3	2
Not at all	1	1

Headteacher intervention COVID-19 survey question 36, n = 15. 'Please indicate the extent to which you consider the following factors to have been barriers or obstacles to Year 6 pupils engaging with the additional literacy support and resources provided since the return to schools in September 2020.'

Figure 24b: Control



	The implementation of social distancing guidelines in school Competing priorities for pupils (e.g. curriculum catch-up)	Student absenteeism due to COVID-19
Don't know	0	0
To a great extent	0	1
Somewhat	1	1
Very little	2	1
Not at all	1	1

Headteacher control COVID-19 survey question 32, n=12. 'Please indicate the extent to which you consider the following factors to have been barriers or obstacles to Year 6 pupils engaging with the additional literacy support and resources provided since the return to schools in September 2020.'

Cost

Using data from the IPE surveys with headteachers and teachers as well as information from the PALS-UK delivery team, the cost evaluation shows that:

- the total estimated cumulative cost of delivering PALS-UK over a three-year period is £181 per classroom. This equates to a cost of £2.57 per pupil per year over a three-year period; this includes the cost of teacher training and resources;
- analysis of the IPE surveys highlighted that schools generally did not hire new staff but some schools indicated that they provided supply cover as a result of teachers participating in the training; and
- some schools indicated that they incurred additional costs in the process of implementing PALS-UK.

This section estimates average costs associated with PALS-UK implementation. We present average marginal costs per pupil over three years following EEF guidelines on cost evaluation. As specified in the Method section, cost data was provided by the delivery team and supplemented by information collected in the headteacher survey at endline.

As reported in the Implementation and process evaluation section of this report, the main programme costs relate to:

- programme fees—for school access to training and materials based on the market value;
- prerequisite costs—any resources already available in the setting that were required to implement PALS-UK;
- staff time for teacher training, preparation, and delivery of PALS-UK, identifying separately the cost of new hires, supply staff, and any extensions made to the contract hours of Year 6 teachers; and
- any additional (unpaid) staff time supporting the delivery of PALS-UK as reported by headteachers.

The estimate of the costs of delivering PALS-UK per pupil per school is based on an average of 24.5 pupils per class in trial schools; the per-pupil cost would clearly be less for larger classes. We assume that PALS-UK sessions replace other time in the classroom and that it does not require paying teachers for more time. The calculation of the cumulative cost also assumes that teachers trained in the programme approach stay at the school for three years (that is, there is no staff change over) meaning that training and material costs apply to the first year of the programme. On this basis, we estimate the cost of implementing PALS-UK over three years for one-form schools assuming that only one teacher participates in the training. The overall cost for a larger school will be higher: for example, a school with two Year 5 classes will need to spend around double (assuming there is no discount available for multi-entry schools). Some savings can occur from sharing some of the resources required for delivery (such as books or IT equipment) but we expect these savings to be small.

The EEF's cost evaluation guidance (EEF 2019) recommends applying a 3.5% discount rate to account for value of money over time, however, given that there are no recurring costs we did not think it necessary We did account for inflation, applying a GDP price deflator index obtained from the Office for National Statistics.[37]

The main cost relates to training, and materials. The intervention resources developed by Nottingham Trent University and Coventry University are available to purchase from Vanderbilt University at £31 per class. The training provided by the delivery team costs £150 per teacher. The information on cost related to teacher training was provided by the delivery team. Table 43 provides a breakdown of the costs associated with the different components of the programme.

The cost-associated questions in the surveys and interviews were focused on assessing any prerequisite costs (such as those related to training or additional materials required for delivery) and any direct and marginal costs directly attributable to a school's participation in the intervention (for example, staff time).

[37] The time series are available on: **https://www.ons.gov.uk/economy/grossdomesticproductgdp/timeseries/mnf3/qna**

Table 43: Cumulative costs of PALS-UK (assuming delivery over three years)

	Nominal values			Cost in analysis year
	Year 1	Year 2	Year 3	£PV in Year 3 prices
PALS-UK (one-form school)	£181	0	0	£188.68
Average number of pupils per classroom-per-school-year				24.5
Cost per pupil-school-year				£2.57

The total estimated cumulative cost (in real values) of delivering PALS-UK over a three-year period in one classroom is £188.68 (Table 43). This equates to £2.57 per pupil per year over a three-year period. This estimate is based on the cost linked to training and materials as these costs have been incurred by all schools. It is important to note that this is the minimum cost associated with delivery. The associated programme fee will be higher for multiple-form entry schools as they will need to cover expenses for more teachers. The core estimate is based on the assumption that schools do not incur additional costs or do not require additional resources to deliver PALS-UK. More than half of the headteachers in the schools that implemented PALS-UK did not report that they needed additional resources for implementing the programme (16 of 26) or they were not sure if the school incurred any additional costs in relation to implementation (3 of 26). Around a quarter of respondents (26.92%, 7 of 26) reported that they needed additional resources—books, computers, iPads—in preparation for PALS-UK. It should be noted here that computers and iPads are not necessarily required to deliver PALS-UK. When additional costs not incurred by all schools are included in the estimate, the cumulative cost of delivering PALS-UK over a three-year period is £1,115.3 per classroom in Year 3 prices or £15.17 per pupil per year (Table 45). According to the EEFs cost rating (see Appendix A: EEF cost rating), even with the additional costs this would still be considered a very low-cost intervention.

Table 44: Cost of delivering PALS-UK, including costs incurred by all schools and additional costs incurred by some schools

Item	Type of cost	Mean cost	Total cost over 3 years	Total costs per pupil per year over 3 years in Year 3 prices	Total cost per pupil per year over 3 years
	Costs incurred by all schools				
PALS-UK training course fee (includes costs for initial and top-up training)	Start-up costs per teacher	£150 (incl. VAT)	£150	£156.37	(£156.37/24.5/3) ≈ £2.13
PALS-UK Resources (e.g., manual)	Start-up cost per setting	£31 (excludes postage)	£31	£32.32	(£32.32/24.5/3) ≈ £0.44
Total costs incurred by all schools			£181	£188.68	(£188.68/24.5/3) ≈ £2.57

Table 45: Cost of delivering PALS-UK, including costs incurred by all schools and additional costs incurred by some school

Additional costs (not incurred by all schools)						
	Type of cost	Mean cost reported	Units needed per classroom	Price per unit required	Total cost per pupil per year over 3	Total cost per pupil per year over 3 years (In Year 3 prices)

		(nominal values)			years (Nominal values)	
(Optional) additional resources: computers or iPads	Pre-requisite cost per school	£500 (£400 - £500+)	1 computer or 1 iPad, 1.5 hours per week	£100[38]	(£100/24.5/3) ≈ £1.36	(£104.24/24.5/3) ≈ £1.41
Additional resources: books	Pre-requisite cost per school	£288.89 (£100-£500)	36.11	£8 per book	(£288.89/24.5/3) = £3.93	(£301.15/24.5/3) = £4.1
Supply cover for staff to attending PALS-UK training	Direct costs of staff time	£400	2 days	£200	(£400/24.5/3) ≈ £5.44	(£416.98/24.5/3) ≈ £5.67
Other provisions for staff to attend PALS-UK training activities (e.g., accommodation, meal allowance)	Direct costs of staff time	£100.00	1	£100	(£100/24.5/3) ≈ £1.36	(£104.24/24.5/3) ≈ £1.42
Total addition costs (average) incurred by some schools	Additional pre-requisite and direct cost	£888.89			(£888.89/24.5/3) ≈ £12.09	(£926.62/24.5/3) ≈ £12.61
Total costs incurred by all school		£181			(£181/24.5/3) ≈ £2.42	(£188.68/24.5/3) ≈ £2.57
Total costs average incurred by some schools plus those incurred by all schools		£1,069.89			(1,069.89/24.5/3) ≈ £14.55	(1,115.3/24.5/3) ≈ £15.17

Staff cost

Staff spent additional time in training, preparing for delivery, or offering support in the form of peer observations. Time devoted to training, teacher cover, preparation, and delivery of PALS-UK is presented in Table 46. Given that PALS-UK was implemented during school hours, we do not consider that time devoted to preparation and delivery would entail additional costs. However, some schools will need to pay for teacher cover to enable school staff to take part in PALS-UK training. Therefore, we consider teacher cover for staff attending PALS training as additional cost and part of the marginal cost calculation (see Table 45).

To ensure staff have the knowledge and skills to deliver PALS-UK sessions, each Year 5 teacher is required to spend 1.5 days in training, or 12 hours.

Table 46: Total time devoted by personnel for training, teacher cover, preparation, and delivery

[38] Computers or tablets are expected to be used for 1.5 hours per week for PALS purposes or for approximately 20% of all uses.

		Year 1 Mean number of hours (min, max)	Year 2 Mean number of hours (min, max)	Year 3 Mean number of hours (min, max)
Training	Year 5 Teacher	12 (12, 12)	0	0
Teacher cover	Year 5 Teacher	12 (0,16)	0	0
Preparation	Year 5 Teacher	74.5 (30, 100)	0	0
	Year 5 TA	28 (5, 60)	0	0
	Headteacher	10 (5, 20)	0	0
Delivery	Year 5 Teacher	30 (26.4, 33)	0	0

Schools were advised to replace guided reading time with PALS-UK sessions and to include the sessions in the existing timetable thus not affecting working hours. However, additional teacher preparation time over a period of 20 weeks (above normal workload) is needed. In the endline surveys, all headteachers agreed that teachers spent some additional time preparing for PALS-UK, but most of them were unable to provide an estimate (66.67%, 14 of 21): as shown in Table 46, responses to this question were wide ranging. A further aspect related to time and cost is the increased workload to other school staff resulting from the PALS-UK programme. Besides teachers responsible for delivery, other school staff, including Year 5 TAs and headteachers, may need to spend additional time supporting delivery. The majority of headteachers indicated that TAs also spent some time supporting teachers (85.7%, 18 of 21). Similarly, 20 of 22 (90.9%) headteachers confirmed that the implementation of PALS-UK did have an effect on their workload. The range of hours school staff spent in supporting PALS-UK activities is reported in Table 46. The main message from these responses is that schools spent additional time providing support for PALS-UK activities. Taking into consideration time spent in training, preparation, and delivery, one Year 5 teacher is required to spend between 42 and 145 hours to deliver PALS-UK over 20 weeks. This is based on the assumption that four weeks of PALS-UK are delivered per month.

Conclusion

Table 47: Key conclusions

Key conclusions
1. Children in PALS-UK schools made the equivalent of one month less progress in reading, on average, compared to children in other schools. This result has a low EEF padlock security rating. Given the impacts of the COVID-19 pandemic on the trial, there is uncertainty about whether this trial can provide a generalisable estimate of the impact of PALS-UK.
2. Children in PALS-UK schools who were eligible for free school meals made the equivalent of one months' additional progress in reading, on average, compared to children in the control group eligible for free school meals. This result has less security than the overall findings and is similarly affected by the impacts of COVID-19 on the trial and the low overall security rating.
3. There was no evidence that PALS-UK impacted pupils' reading self-efficacy, based on pupil responses to a self-efficacy questionnaire.
4. Teachers surveyed and interviewed as part of the process evaluation felt that PALS-UK had positive effects on pupils' reading confidence and reading fluency and, to a lesser extent, on reading attainment.
5. Observations of programme delivery suggest that most schools implemented PALS-UK with excellent fidelity. Survey data also indicates that 100% of schools delivered PALS-UK for the full 20-week period and over 95% of teachers attended the initial and top-up training, although low survey response rates could have biased the results. These findings suggest that the programme was delivered well from October 2019 to March 2020, although due to COVID-19 disruption, schools could not continue to deliver PALS-UK as intended between March 2020 and the delayed start to outcome testing in November and December 2020.

Impact evaluation and IPE integration

In this evaluation we undertook a cluster randomised controlled design to estimate the effect of PALS-UK on pupil outcomes as outlined in the logic model (Figure 1). This trial included 89 schools and had an overall attrition rate of 25.8% at the school level and 36.1% at the pupil level. Balance was achieved on the pre-specified characteristics and the trial was powered to detect effect sizes of 0.287 with a power level of 80%.

The attrition level is too high and unbalanced across groups (that is, the attrition rate is 45.1% in the control group while in the intervention group it is 27.5%) to allow us to make robust inferences. We also acknowledge that evidence on impact is limited with rescoped trial design due to the COVID-19 pandemic (see

Methods section for more information on trial adaptations due to COVID-19). Similarly, the findings in the IPE build upon evidence provided only by those schools that engaged in the surveys or interviews. A higher response rate on the surveys would have allowed for more definitive conclusions. A comprehensive discussion on the limitations of this study is presented below in Limitations and lessons learned.

The primary outcome impact results found no conclusive evidence of any positive or negative impact of PALS-UK on reading attainment. While the best estimate in this evaluation found an effect equivalent to one month less progress made by pupils in the PALS-UK group compared to those in the control group, there was a high level of statistical uncertainty in the results. No conclusive positive or negative effect of PALS-UK was found on the secondary outcome (self-efficacy in reading survey).

The evaluation found no conclusive evidence of any positive or negative impact of increased compliance with the PALS-UK intervention due to the high level of statistical uncertainty in the results. While the main compliance model pointed towards a small positive effect on reading attainment, the magnitude was small (equivalent to zero months' additional progress) and the difference was not statistically significant. These findings should also be interpreted with caution as they are based on a small sample of schools that were able to supply data used for compliance.

Further, the evaluation did not find any conclusive evidence of differential impacts of PALS-UK on the reading attainment of the pupil subgroups of interest in this study (FSM, EAL, and prior reading attainment), again due to the high level of statistical uncertainty in the results. It should be noted that the subgroup analyses were limited by power issues related to limited sample sizes due to attrition and again the difference was not significant at traditional significance levels. The best estimate in this trial suggests that pupils eligible for FSM made one additional month's progress in reading while EAL pupils made zero months' additional progress compared to the control group. However, these results are unable to provide conclusive evidence due to the high degree of statistical uncertainty associated with them. Similarly, the analysis of the impact of PALS-UK on the reading attainment of pupil subgroups with varying levels of reading attainment at baseline were inconclusive. While pupils in PALS-UK schools with lower (one month less progress), very low (two months less progress), higher (one months' less progress in the PALS-UK group) and very high (zero months' additional progress in the PALS-UK group) baseline reading attainment achieved lower scores in endline reading attainment, on average, than their counterparts in the control group, the mean difference in scores was small and there was a high level of statistical uncertainty around these results.

There are several potential interpretations of the inconclusive findings around the impact of PALS-UK in this trial. Taking into consideration the evidence from IPE data (surveys and interviews), one possibility is that the effect of the intervention became diluted over time. Despite its shortcomings, data from the surveys and interviews with teachers and headteachers in participating schools undertaken shortly after programme delivery suggests that the programme had a positive impact on pupils immediately after delivery, whilst data from the survey and interviews collected in the next academic year suggests that the positive effect was not sustained and the pandemic had negatively impacted on pupils' literacy development going into Year 6.

One mechanism by which the impact of PALS-UK may have been diluted was that schools have not continued to employ the PALS-UK approach following school closures and the move to remote learning and that, as a result, any early programme effects were not detectable in the follow- up period. Indeed, findings from the IPE data collection in the COVID-19 phase found that just less than half of teachers in intervention schools reported delivering PALS-UK to Year 6 pupils mainly due to limited capacity and competing priorities. The IPE data suggests that it was challenging for schools to start incorporating the programme in their usual timetable and sustain this or to offer structured support to pupils in a COVID-19 affected environment. Considering the wider context, it is important to flag that the COVID-related move to remote learning (including literacy learning) and social distancing for a five-month period immediately following the delivery of PALS-UK may have been a main key attributing factor to stagnation in reading development. The negative effect of COVID-19 may have diluted any positive effect of the PALS-UK intervention over time. However, it is difficult to draw any robust inferences on this potential dilution effect due to the high and unbalanced attrition rates in the PALS-UK and control groups, as mentioned above.

Another consideration for interpretation is that measures focusing on reading fluency are more sensitive to detect effects of this intervention in comparison to measures that monitor change in reading attainment, that is, the PiRA test scores used for the primary outcomes analyses. Indeed, as specified in the programme logic model, programme activities should first produce noticeable improvements in reading fluency and comprehension, which then ultimately result in improvements in reading attainment in the longer term. Despite the initial plan to measure reading fluency and

comprehension, it was not feasible to complete this testing element due to the COVID-19 outbreak in January 2021. As evidenced from the IPE data, school staff felt that pupils who implemented the PALS-UK approach were making positive improvements in reading fluency. So while pupils exposed to the PALS-UK programme in this trial may have improved their reading fluency and comprehension noticeably, this programme effect could not be detected in a robust way.

A final consideration, as documented in the Usual Practice section of the IPE, is that a considerable proportion of control schools at baseline and, to a lesser extent at endline, reported implementing other reading programmes in their school, the most commonly mentioned being Accelerated Reader. Thus, it is possible that the implementation of these kinds of programmes in control schools may dilute any effect of PALS-UK on pupils in intervention schools.

Findings from the IPE at a high level can be summarised as follows. Compliance with programme activities was generally very high in relation to attendance at the initial full-day and half-day top-up training sessions, the delivery of the four-week pupil training, and delivery of the 16-week intervention.

With regards to implementation fidelity, information received about PALS-UK prior to implementation was generally deemed clear, although reported levels of preparedness among intervention schools at baseline were varied. The training programme and supporting resources were generally deemed high quality and useful for implementation by teachers, although there were some areas of contention here—for example, survey tasks and email reminders were not deemed so useful and some teachers questioned the usefulness of pupil training for informing pupils how to work well in pairs. Schools generally seemed to be satisfied with the ongoing support provided by the delivery team as well as wider support for delivery within their school. Finally, most schools indicated that they would continue with PALS-UK beyond the end of the trial but many mentioned conditions, such as reducing the number of sessions per week.

Findings on the perceived impact of PALS-UK helps to interpret the inconclusive results yielded from the impact evaluation. Indeed, while teachers' views in the endline survey suggest that PALS-UK had improved reading attainment, their views based on the follow-up survey during the COVID-19 driven phase suggest that any potential gains in reading linked to the use of PALS-UK had been lost by the end of the trial due to the pandemic. However, it is important to note that interview responses indicated that school staff were more positive about programme impacts on reading fluency and comprehension, which could not be measured at endline due to the COVID-19 pandemic. Furthermore, school staff mentioned some other outcomes of PALS-UK not captured in the programme logic model, including improved pupil vocabulary and learning independence.

With regards to usual practice, at baseline the majority of schools across intervention and control arms had dedicated guided reading time and just under half were implementing a reading programme, the most commonly mentioned being Accelerated Reader. At endline, while the majority of schools across intervention and control still had dedicated guided reading time, only a minority were implementing a reading programme, the most common still being Accelerated Reader. Finally, during the COVID-19-driven phase, while the majority of intervention schools confirmed that they were providing additional support and resources for improving reading fluency and comprehension to Year 6 pupils, the evidence was mixed in control schools.

The total estimated cumulative cost of delivering PALS-UK over a three-year period is £181 per classroom. This equates to a cost of £2.57 per pupil per year over a three-year period. This cost comprises the cost for training (for one teacher) and resources. Analysis of the IPE surveys has highlighted that schools generally did not hire new staff but some indicated that they provided supply cover during teacher training; some indicated that they incurred additional costs in the process of embedding PALS-UK in their school, such as purchasing reading books or computer/tablet devices.

Our key finding, therefore, is that due to statistical uncertainty there is no conclusive evidence of any positive or negative impact of PALS-UK on reading attainment, although this evaluation found a small, negative overall effect equivalent to one month less progress for pupils in the PALS-UK group. Secondly, the process evaluation established that there are no significant concerns around implementation, that is, compliance with the mandatory programme element was high and the programme was implemented with high fidelity. Thirdly, teachers perceived PALS-UK positively, including the training and the structured PALS-UK sessions, and they felt that the programme had positive effect on pupils' attainment, and particularly on reading fluency. Last, IPE data suggests that the COVID-19 pandemic may have been the main contributing factor for reading stagnation and may have diluted any positive effect of PALS-UK in the medium term.

Evidence to support the logic model

Given the delay in data collection and considering the wider COVID context, the current evaluation is not able to draw firm conclusions on how well the proposed logic model works in practice, that is, whether PALS-UK changes reading attainment. As mentioned above, a key missing link here is whether exposure to the programme first results in improvements in reading fluency and comprehension, which were not measured in this trial due to the COVID-19 situation. Given that outcome testing was delayed and considering that schools in this trial were affected by the COVID-19 pandemic it is difficult to draw firm conclusions and to assess what would have been the impact of PALS-UK in a 'normal' environment and what would have been the impact of PALS-UK in the short run.

The IPE data suggests that the key programme inputs, activities, and outputs as delivered in the trial are acceptable to schools and they were implemented in the short term with high fidelity. However, the response rates in the surveys at endline and during the COVID-19 driven phase were relatively low and, as a consequence, do not provide comprehensive insight into compliance and implementation fidelity. Nonetheless, given the inconclusive findings from the impact evaluation and the high level of compliance and fidelity with PALS-UK programme activities among schools as identified in the IPE, the evaluation did not yield any rationale to update or adapt the logic model as shown in Figure 1.

Limitations and lessons learned

Attrition and level of missingness in the outcome data and rescoped outcome testing due to COVID-19 restrictions

From a research design perspective, the COVID-19 restrictions created significant disruptions in organising outcome testing as well as in collecting IPE data. As a result, one of the most significant issues that limits the extent to which we can draw firm conclusions is the extent of the missing outcome data on the primary and secondary outcome measures due to the challenges encountered in organising testing due to COVID-19 related constraints. Indeed, findings from the missing data analysis show that the rate of missingness in the primary outcome at endline (PiRA) was as high as 36.1% due to schools being unable to participate in the outcome testing due to COVID-19 related challenges. The result of this high rate of attrition was a noticeable increase in the MDES for the overall sample from 0.23 at randomisation to 0.29 at primary analysis stage. Furthermore, attrition disproportionately affected pupils in the control group as well as pupils with lower reading attainment at baseline, further reducing the extent to which firm conclusions on the impact of PALS-UK on reading attainment can be drawn.

The initial trial was designed to test the immediate effect of PALS-UK. There were initial plans in place to undertake outcome testing in May and June 2020 but all parties agreed that it was not feasible due to the COVID-19 situation. This meant that outcome testing began eight months after delivery finished in which time pupils may have had varied reading and wider learning experiences, which meant that this evaluation tested the impact of PALS-UK in the medium term in a COVID-19 affected environment.

Furthermore, only 18 schools were able to participate in outcome testing measuring reading comprehension before it was decided to forego secondary outcome testing aligned with reading fluency and comprehension due to COVID-19.

Secondary outcome measures—oral reading fluency and reading comprehension

As we have described in the report, according to the programme logic model positive change in reading development associated with usage of the PALS-UK approach would be first observed in reading fluency and comprehension rather than in reading attainment. This trial was not designed to detect any differences in oral reading fluency and reading comprehension as the initial plan was to administer the measure to a subsample of pupils. Considering the underlying theory, and the IPE evidence that is suggestive of positive change in oral reading fluency, we recommend a future evaluation to be powered to capture such impacts in a robust way. We do not have any concerns about the measures in the WIAT-III-UK-T Oral Reading Fluency and Reading Comprehension subtests.

Subgroup analysis

This trial was not powered to provide a precise estimate for the impact of PALS-UK on the reading attainment for the subgroups of interest—FSM pupils, EAL pupils, and pupils with low or very low reading attainment. Furthermore, assessing the perceived impact on these subgroups through the IPE was out of the scope of this evaluation. This

suggests that a future study that is powered to detect subgroup differences is needed so the potential effect of the intervention on these subgroups can be captured with more certainty.

Limitations related to the IPE

A final limitation was the limited response to the IPE surveys meaning that the IPE findings build upon evidence provided only by those schools and staff that engaged in the surveys. This was particularly problematic in the follow-up surveys in the COVID-19 phase where the achieved response rate was approximately 40% to 48% in the treatment schools and the response rate in control schools was less than 18%. While response rates were stronger at endline (approximately 63% to 91% in intervention schools, 71% in control schools), non-responses to this degree still made it more difficult to draw firm conclusions about compliance and implementation fidelity. While data from responsive schools indicates that compliance and fidelity to intended programme implementation was high, it is difficult to make an assessment on the extent to which schools perhaps less engaged with the programme were not captured in the IPE data collection.

Also, we would have, if possible, followed the same case study schools in their journey from the conclusion of the delivery period (April 2020) to outcome testing (November to December 2020). This would have allowed us to further capture within-school changes between the time delivery ended and outcome testing took place. However, only one of the case study schools that was interviewed in the initial round agreed to participate in the second round and, as a result, we approached schools on our waiting list.

Future research and publications

Given the scope of the evaluation and the challenges presented by the COVID-19 restrictions, we perceive that future research on the PALS-UK programme should focus on the following aspects.

Short-term programme impacts

This evaluation was initially designed to test short-term programme impacts. However, plans to administer outcome testing immediately after the programme delivery period were cancelled due to COVID-19 challenges. Therefore, it would be relevant a future evaluation to explore **short term impact** closer to the 20-weeks of PALS-UK delivery. This is especially important given that findings from the IPE indicated that initial gains in reading progress immediately after programme delivery were somewhat diluted by the COVID-19-driven phase, which is closer to when outcome testing was undertaken.

Impact of PALS-UK on oral reading fluency and reading comprehension

As elaborated in the section on Nonetheless, given the inconclusive findings from the impact evaluation and the high level of compliance and fidelity with PALS-UK programme activities among schools as identified in the IPE, the evaluation did not yield any rationale to update or adapt the logic model as shown in Figure 1.

Limitations and lessons learned, assessing and understanding the PALS-UK effect on oral reading fluency and comprehension seems to be important considering the IPE findings on the perceived programme impact.

Impact of PALS-UK on the reading attainment of pupil subgroups

Also mentioned in the Nonetheless, given the inconclusive findings from the impact evaluation and the high level of compliance and fidelity with PALS-UK programme activities among schools as identified in the IPE, the evaluation did not yield any rationale to update or adapt the logic model as shown in Figure 1.

Limitations and lessons learned section, future research should increase the focus on understanding and assessing programme effect on the specific groups of pupils, including pupils eligible for free school meals, where an effect was detected in this study despite being underpowered.

Impact of PALS-UK on peer relationships

While outside of the scope of this evaluation, a secondary outcome of the PALS-UK programme is improved peer relationships among pupils as a result of participating in paired reading activities. However, findings from the IPE in this study suggested that pairing pupils was challenging in some schools for a number of reasons, including difficulties in

selecting pupils that could work well together and other behavioural issues. Future research should therefore include in its scope an evaluation of whether the PALS-UK approach has any impact on peer relationships and, if so, what are the mechanisms driving the effect.

Finally, given that PALS-UK was well received and that the majority of teachers indicated that they would like to continue delivery if the programme was available in the future, we suggest that there is merit in conducting a re-trial of PALS-UK. However, based on findings from the IPE, we acknowledge that some minor tweaks to elements of the programme may be beneficial for future implementation of the intervention—for example, how to inform pupils on how to work well in pairs in the pupil training, or how to inform teachers to select appropriate reading books for pairs in the top-up training.

References

Allison, P. D. (2003) 'Missing Data Techniques for Structural Equation Modelling', *Journal of Abnormal Psychology*, 112 (4), p. 545.

Allison, P. D. (2001) 'Missing Data' (Quantitative Applications in the Social Sciences series no. 136), Thousand Oaks: Sage.

Angrist, J. D. and Keueger, A. B. (1991) 'Does Compulsory School Attendance Affect Schooling and Earnings?', *Quarterly Journal of Economics*, 106 (4), pp. 979–1014.

Angrist, J. D. (2006) 'Instrumental Variables Methods in Experimental Criminological Research: What, Why and How', *Journal of Experimental Criminology*, *2* (1), pp. 23-44.

De Boer, M.R., Waterlander, W.E., Kuijper, L.D., Steenhuis, I.H. and Twisk, J.W. (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*(1), pp.1-8.

Burns, T. G. (2010) 'Wechsler Individual Achievement Test-III: What is the "Gold Standard" for Measuring Academic Achievement?', *Applied Neuropsychology*, 17 (3), pp. 234–236.

Carpenter, J. and Kenward, M. (2012) *Multiple Imputation and Its Application*, Chichester: John Wiley.

Carril, A. (2017) 'Dealing with Misfits in Random Treatment Assignment', *Stata Journal*, 17 (3), pp. 652–667.

Carroll, J. M. and Fox, A. C. (2017) 'Reading Self-Efficacy Predicts Word Reading But Not Comprehension in Both Girls and Boys', *Frontiers in Psychology*, 7, p. 2056.

Dimova, S. and Ilie, S. (2021) 'Peer Assisted Learning Strategies for Reading UK (PALS-UK) Statistical Analysis Plan' (updated in response to the COVID-19 pandemic): **https://educationendowmentfoundation.org.uk/public/files/Projects/PALS_SAP_update_2021.pdf**

Dimova, S., and Sutherland, A. (2019) 'Peer Assisted Learning Strategies for Reading UK (PALS-UK) Evaluation Protocol': **https://d2tic4wvo1iusb.cloudfront.net/documents/projects/EEF_trial_protocol_PALS_-UK_v_0102.pdf?v=1630925324**

Dimova, S., Sutherland, A. and Culora, A. (2021) 'Peer Assisted Learning Strategies for Reading UK (PALS-UK) Evaluation Protocol' (updated in response to the COVID-19 pandemic): **https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Protocols/PALS_trial_protocol_update_2021.pdf**

Dong, N. and Maynard, R. (2013) 'PowerUp!: A Tool for Calculating Minimum Detectable Effect Sizes and Minimum Required Sample Sizes for Experimental and Quasiexperimental Design Studies', *Journal of Research on Educational Effectiveness*, 6, pp. 24–67. DOI: 10.1080/19345747.2012.673143

EEF (2018a) 'The EEF Teaching and Evidence Toolkit', London: Education Endowment Foundation: **https://educationendowmentfoundation.org.uk/evidence-summaries/teaching-learning-toolkit/peer-tutoring/**

EEF (2018b) 'Statistical Analysis Guidance for EEF Evaluations', London: Education Endowment Foundation: https://educationendowmentfoundation.org.uk/public/files/Grantee_guide_and_EEF_policies/Evaluation/Writing_a_Protocol_or_SAP/EEF_statistical_analysis_guidance_2018.pdf

EEF (2019) 'EEF Guidance on Cost Evaluations', London: Education Endowment Foundation: https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting_up_an_Evaluation/Cost_Evaluation_Guidance_2019.12.11.pdf

Fielding, S., Ogbuagu, A., Sivasubramaniam, S., MacLennan, G. and Ramsay, C. R. (2016) 'Reporting and Dealing with Missing Quality of Life Data in RCTs: Has the Picture Changed in the Last Decade?', *Quality of Life Research*, 25 (12), pp. 2977–983.

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).

Higgins, J., Altman, D., Gøtzsche, P., Jüni, P., Moher, D., Oxman, A., Savovic, J., Schulz, K., Weeks, l. and Sterne, J. (Cochrane Bias Methods Group, Cochrane Statistical Methods Group) (2011) 'The Cochrane Collaboration's Tool for Assessing Risk of Bias in Randomised Trials', *BMJ*, 343, d5928.

Hox, J. (1998) 'Multilevel Modeling: When and Why', in Balderjahn, I., Mathar, R. and Schader M. (eds), *Classification, Data Analysis, and Data Highways* (Studies in Classification, Data Analysis, and Knowledge Organization series), Berlin, Heidelberg: Springer.

Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R. and Kerr, K. (2016) 'Implementation and Process Evaluation (IPE) for Interventions in Education Settings: An Introductory Handbook', London: Education Endowment Foundation.

Lloyd, C., Edovald, T., Kiss, Z., Morris, S., Skipp, A. and Ahmed, H. (2015) 'Paired Reading: Evaluation Report and Executive Summary', London: Education Endowment Foundation: 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: https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/Shared_Maths.pdf

Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B. and d' Apolonia, S. (1996) 'Within-Class Grouping: A Meta-Analysis', *Review of Educational Research*, 66, pp. 423–458.

McCarty, C. and Ruttle, K. (2018) *Progress in Reading Assessment Manual (Stage 2)*, Second Edition, Hodder Education.

Pearson Clinical. (n.d). 'Description of the WIAT-III Subtests': **https://www.pearsonclinical.co.uk/Sitedownloads/Productpdfs/wiat-iii-uk-subtests-description.pdf**

Roberts, C. and Torgerson, D. (1999) 'Baseline Imbalance in Randomised Controlled Trials', *BMJ*, 319:185.

Sáenz, L. M., Fuchs, L. S. and Fuchs, D. (2005) 'Peer-Assisted Learning Strategies for English Language Learners with Learning Disabilities', *Exceptional Children*, 71 (3), pp. 231–247.

Senn, S. (1994) 'Testing for baseline balance in clinical trials', *Statistics in Medicine*, 13 (17), pp. 1715 - 26.

Shenderovich, Y., Thurston, A. and Miller, S. (2016) 'Cross-Age Tutoring in Kindergarten and Elementary School Settings: A systematic Review and Meta-Analysis', *International Journal of Educational Research*, 76, pp. 190 - 210.

Snijders, T. A. and Bosker, R. J. (1994) 'Modeled Variance in Two-Level Models', *Sociological Methods and Research*, 22 (3), pp. 342–363.

Snijders, T. A. (2005) 'Power and Sample Size in Multilevel Modeling', *Encyclopedia of Statistics in Behavioral Science*, 3 (157), p. 1573.

Spybrook, J., Kelcey, B. and Dong, N. (2016) 'Power for Detecting Treatment by Moderator Effects in Two- and Three-Level Cluster Randomized Trials', *Journal of Educational and Behavioral Statistics*, 41 (6), pp. 605–627. https://doi.org/10.3102/107699861665544290-210

Strand, S. and Hessel, A. (2018) 'English as an Additional Language, Proficiency in English and Pupils' Educational Achievement: An Analysis of Local Authority Data'. Cambridge: Bell Foundation: https://www.bell-foundation.org.uk/app/uploads/2018/10/EAL-PIE-and-Educational-Achievement-Report-2018-FV.pdf

Sullivan, T. R., White, I. R., Salter, A. B., Ryan, P. and Lee, K. J. (2018) 'Should Multiple Imputation Be the Method of Choice for Handling Missing Data in Randomized Trials?', *Statistical Methods in Medical Research*, 27 (9), pp. 2610–626.

Thomas, D. R. (2006) 'A General Inductive Approach for Analyzing Qualitative Evaluation Data', *American Journal of Evaluation*, 27 (2), pp. 237–246.

Topping, K., Millder, D., Thurston, A., McGavock, K. and Conlin, N. (2011) 'Peer Tutoring in Reading in Scotland: Thinking Big', *Literacy*, 45 (1), pp. 3–9.

Tymms, P., Merrell, C., Thurstone, A., Andor, J., Topping, K. and 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), pp. 265–289.

White, I. (2018) 'RCTMISS: Stata Module to Analyse a Randomised Controlled Trial (RCT) Allowing for Informatively Missing Outcome Data': https://ideas.repec.org/c/boc/bocode/s458304.html

WWC (2012) 'Peer-Assisted Learning/Literacy Strategies', What Works Clearinghouse, Institute of Education Sciences, U.S. Department of Education: https://files.eric.ed.gov/fulltext/ED531596.pdf

Appendix A: EEF cost rating

Figure 25: 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

OUTCOME: *Reading Attainment (PiRA, 1-40 (age standardised <70–>130), designed by Rising Stars)*

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%			Adjustment for threats to internal validity [-1]		
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%					
2	Design for comparison that considers selection only on some relevant confounders	0.40 - 0.49	31-40%	2				
1	Design for comparison that does not consider selection on any relevant confounders	0.50 - 0.59	41-50%					1
0	No comparator	>=0.6	>50%					

Threats to validity	Threat to internal validity?	Comments
Threat 1: Confounding	Low	Well-designed RCT, with randomization by evaluation team. Controlled for pre-test measure. No additional risks posed beyond the risks already captured in the high degree of loss to follow up.
Threat 2: Concurrent Interventions	Moderate	Evidence from IPE of some schools taking up reading initiatives (e.g. accelerated reader) in the control arm.
Threat 3: Experimental effects	Low	Due to circumstances, tests were administered by Y6 teachers. However, marking was undertaken independently.
Threat 4: Implementation fidelity	Moderate	Implementation was reportedly good, although survey response rates in IPE were very low. Direction of any potential bias difficult to identify.
Threat 5: Missing Data	High	Considerable missing data influenced by COVID-19 disruption to schools. Differential attrition, higher among control group schools. Missing data analysis conducted; uncertainty in accounting for potential biases.
Threat 6: Measurement of Outcomes	High	Delayed outcome testing influenced by COVID-19 related school closures
Threat 7: Selective reporting	Low	Protocol and SAPs published; trial registered. Evaluation followed protocol closely. Secondary outcome testing commenced but had to be terminated.

- **Initial padlock score:** 2 - Padlocks – Well-designed two-arm cluster randomized trial; MDES of .231 at randomization. Circumstances led to high experienced attrition of 36.1%.
- **Reason for adjustment for threats to validity**: 1 Padlock lost - Although the presence of Accelerated Reader in the control arm is a challenge for the trial, it is not a deviation from the counterfactual that would have been observed in the intervention arm had the arms been reversed. Some threats are captured in the initial padlocks related to loss to follow-up. However delayed outcome testing by 6 months is significant and high-risk threat to validity of measured outcomes along with the differential attrition experienced in the trial. There is a considerable threat to the external validity – COVID – but this is not captured in the padlocks.
- **Final padlock score:** 1 padlock.

Appendix C: Recruitment Documentation

Original PALS-UK Memorandum of Understanding (MoU)

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

MEMORANDUM OF UNDERSTANDING

This information is for schools wishing to take part in the EEF funded Evaluation of Peer Assisted Learning Strategies (PALS-UK) (https://vkc.mc.vanderbilt.edu/frg/about-3/) research project in the 2019/20 and 2020/21 academic years. Please read this information carefully.

Aims of the Evaluation

The aim of this project is to evaluate the effects of PALS-UK on reading comprehension and fluency for pupils in Year 5. Existing research suggests that PALS can benefit schools in multiple ways, such as raising the profile of language in primary schools and providing schools with professional support for teaching and learning. Furthermore, studies have found positive effects of peer tutoring in relation to reading skills, such as accuracy, comprehension and fluency. By participating in this evaluation, schools and individual participants contribute to furthering our knowledge on strategies for improving pupil's reading fluency and comprehension.

The evaluation is funded by the Education Endowment Foundation (EEF). The intervention PALS-UK is being delivered by Coventry University (the Delivery team) and will be independently evaluated by RAND Europe (the Evaluation team)[1] .

PALS-UK

PALS-UK is a 20-week whole class structured paired reading intervention that aims to improve pupils' reading comprehension and fluency. Pupils work in pairs to complete four activities, three times a week for 30-35 minutes each time. Training is provided to pupils to support them in assisting their peer's reading development. Teachers support implementation by offering encouragement and support where needed. This project is to assess the effectiveness of PALS-UK, which is only available by this EEF funded trial.

Structure of the Evaluation

The evaluation of PALS-UK comprises a 'randomised controlled trial' (RCT). This means that half the schools, will be randomised to receive the intervention (the PALS-UK intervention). These are the Intervention Schools. The other half will not receive the intervention. These are the Control Schools. Once the PALS-UK intervention is completed, outcomes from children in the Intervention Schools will be compared to those in the Control Schools to find out whether the intervention has made a measurable difference. The random allocation of schools is essential to the evaluation, as it is the

best and most rigorous way to find out the effects of the PALS-UK intervention on children's reading attainment. It is therefore important that schools understand, and consent to, the random allocation process, meaning that each school may or may not be allocated to PALS-UK intervention, as determined by the randomisation process.

At the initial stage of the research project teachers will administer a reading test, this will be overseen by Coventry University in September 2019. Following the initial assessment, half of the schools that join the study are selected to deliver PALS-UK. Year 5 teachers in the schools selected to deliver PALS-UK ('intervention' school) will undertake training, training dates will be provided.

Teachers will attend one whole day training in September 2019, in the week beginning the 23rd of September, to cover how to deliver PALS-UK, training will take place in both the North East and the Midlands. In addition, four weeks following the initial training, half-day training will be provided by Coventry University in both the North East and the Midlands. Attendance at the PALS-UK training is mandatory for Intervention Schools.

The intervention period will start after the initial training, around October half term. Teachers will spend four weeks training their pupils on the PALS-UK approach and how to be an effective partner, all the materials are provided. These first four weeks are part of the 20-week intervention. Following the training blocks PALS-UK will be implemented for 16 weeks. Schools must commit to allocating 3 x 30-35 minutes sessions a week for 20 weeks to participate in the evaluation. These sessions can replace guided reading or equivalent.

Intervention outcomes will be assessed by examining pupil's outcomes at the end of school year 2019/20. We will measure all pupils overall reading and then a sub-sample of pupil's reading fluency, reading comprehension skills and reading self-efficacy in the end of school year 2019/20. The data collectors will be appropriately trained and have full Disclosure and Barring Service clearance. Furthermore, members of the Evaluation Team may be present to observe some of thetraining sessions as part of the independent evaluation.

Regardless of whether the school is an intervention or control school, data on Year 5 pupils will be collected to assist with studying the impact of the intervention. We will be asking for the following information: Year 5 pupil's name, gender, date of birth, Free School Meals status and Unique Pupil Number. In order to study longer-term impact pupil data will be linked to national data in 2019/20. RAND Europe will use the information provided to link with the National Pupil Database (which is held by the Department for Education) to collect pupil's Key Stage 2 results in English.

Costs

Intervention schools will be **charged a fee of £200** to take part in the PALS-UK intervention, to cover teacher training and resources including books and the intervention manual. Control Schools will **receive a payment of £500 per school** on completion of the trial. They may spend this on PALS-UK once the evaluation is completed at the end of 2020/21.

Roles and Responsibilities

The Delivery Team (Coventry University) will:

- Be the first point of contact for any questions about the PALS-UK intervention.
- Collate names, gender, date of birth, and Unique Pupil Numbers of Y5 pupils received from schools, and send this to the Evaluation Team.
- Oversee the pre-testing in September 2019.
- Ensure that teachers have freely given informed consent to take part in the research.

- Provide PALS-UK initial and top-up training to Year 5 teachers and peer observers.
- Provide feedback and support to the Y5 teachers via e-mail and telephone support delivering PALS-UK.
- Conduct observations of some of the PALS-UK sessions to support delivery.
- Communicate with the schools about all aspects of PALS-UK training, delivery and assessment.
- Support publication and dissemination of evaluation findings at the end of the study.

The Evaluation Team (RAND Europe) will:

- Be the first point of contact for any evaluation-related questions about the trial.
- Design and conduct the independent evaluation of PALS-UK, which will include the following activities:
    - Randomization of participating school to treatment/control condition (Autumn term 2019/20).
    - Development and implementation of data collection tools- including online surveys of head teachers, Year 5 teachers and qualitative paper survey with pupils, and, in a small number of schools, observations of PALS training and interviews with head teachers and Year 5 teachers.
    - Administer surveys to head teachers at the beginning of the study (Autumn term 2019/20).
    - Administer surveys to head teachers and Year 5 teachers at the end of the study (Summer term 2019/20).
    - Administer primary outcome assessments (PiRA reading assessment and self-efficacy scale) with Year 5 pupils at the end of the study (Summer Term 2019/2020).
    - Administer secondary outcome assessments (WIAT-III oral reading fluency and reading comprehension) with at least 1/3rd of Year 5 pupils at the end of the study (Summer Term 2019/2020)
    - Administer qualitative paper survey with pupils at the end of the study (coincides with outcome testing, Summer term 2019/20).
    - Observe training sessions delivered by Prof. McMaster (Autumn term 2019/20).
    - Conduct interviews with head teachers and Y5 teachers in selected schools (Summer term 2019/20).
    - Analyze the outcome and process data from the study.
    - Using the Unique Pupil Numbers of Year 4 pupils received from schools in school year 2018/19 to link to National Pupil database to compile follow up data for school year 2020/2021.
    - Ensure all members of the Evaluation Team are appropriately trained and have full Disclosure and Barring Service (DBS) clearance.
    - Publish a report on the findings of the study and disseminate the evaluation findings at the end of the study.

The Schools will:

- Read the School Information Booklet and Privacy Notice before signing the MOU.
- Ensure that parents and pupils are fully informed about the nature of the research project and their rights to withdrawal of data in line with GDPR (2018) requirements (see section on data security below). The evaluation team (RAND) will provide information sheets and withdrawal forms to send to parents.
- Distribute teacher information sheet and consent/withdrawal forms provided by Coventry University.
- Distribute the parental information sheets and consent forms/withdrawal forms provided by RAND to parents.

- Collect consent forms from parents (opt-out consent) and share with RAND Europe.
- Securely share information on Year 5 pupils including their name, gender, date of birth, and Unique Pupil Number, as well as the name of their Classroom Teacher and the children they are paired with.
- Consent to randomized allocation.
- Complete pre-test assessment in the first two weeks of September. All information will be provided by Coventry University.
- Head teachers and SLTs will set expectations for involvement and project delivery. Provide teachers the time to deliver the intervention
- Release Y5 teachers to attend PALS training between September and October 2019. It is vital that schools support classroom teachers by allowing them to participate in training and giving them sufficient time to prepare for delivering PALS-UK. Classroom teachers need to attend 1 day initial training week beginning the 23rd of September 2019 before beginning the intervention, half day top-up training 4 weeks later, and time to complete several short online activities throughout the year.
- Ensure that trained teachers deliver the PALS-UK intervention for the minimum requirement of 20 weeks starting before the October half-term (dates to be confirmed), with 3 sessions a week lasting 32 minutes a time. Schools may wish to continue for the academic year and this is acceptable, but 20 weeks needs to have been completed in full before post-test assessments. Timetable changes to accommodate the intervention will need to be made to enableclass teachers to deliver the intervention.
- Consent that assessment will be implemented in only one randomly selected Y5 classroom (although PALS-UK could be implemented in more than one classroom).
- Complete the end-point Year 5 reading assessment and fluency, and pupil's self-efficacy survey to be carried out by in the Summer term 2020.
- Accommodate any additional data collection by the Evaluation Team as required - interviews, observations of PALS sessions or document analysis.
- Accommodate the implementation and process evaluation data collection to be carried out by the Evaluation Team at the beginning and end of the study:
    - Head teacher in intervention and control schools participate in short online surveys in the beginning and in the end of year 2019/2020.
    - Year 5 Teacher in intervention and control schools will participate in short online surveys in the end of year 2019/2020.
    - Pupils' answer survey exploring their views on PALS and pupil training
    - Interviews – a selection of teachers and head teachers from intervention schools will be invited for semi-structured in-depth telephone interviews to discuss their experience of the evaluation. Participation will be voluntary.
    - Attendance of training session
    - To conduct the evaluation, schools are asked to consent to sharing the specified data with the delivery team (Coventry University) and the evaluation team (RAND Europe). Schools will notify the delivery team immediately if a school wishes to withdraw from the intervention or the data collection.

No individual child, teacher or staff member will be identified in any reports arising from the research.

Use of data

All data, including children's test responses and any other pupil data, will be treated with total confidentiality. Data will be collected by the Delivery Team and Evaluation Team and analysed by the Evaluation Team. For the purpose of research, the responses will be linked with information about the pupil from the National Pupil Database (NPD) and shared with Coventry University, the Department for Education, the EEF, EEF's data processor for the data archive and, in an anonymised

form, with the Office for National Statistics and potentially other research teams and the UK Data Archive. Further matching to NPD data and other administrative data may take place during subsequent research. Any data collected will be treated with the strictest confidence. No individual setting, parent or child will be identified in any report arising from the research. The study has received full ethical approval from RAND Europe and the Coventry University. Please see the Privacy Notice for Staff, the Privacy Notice for parents and the EEF's privacy statement for further detail on how personal data will be collected and used (these documents are available here: http://redocuments.org/PALS).

Data Collection Milestones

At the beginning of the study (2019):

- In summer 2019, Coventry University will contact ALL schools to gather data about Year 4 children (names, gender, dates of birth and Unique Pupil Numbers) once parents have been sent consent and withdrawal forms detailing the research project and their rights. This will help with the evaluation and enable linkage to the National Pupil Database. If parents choose to opt out from the research or withdraw their children's data to be collected, their data will not be collected.

At the beginning of school year (2019/20):

- Schools will be randomly allocated to receive PALS-UK or to be in the control group.
- Year 5 teacher and peer observer in participating schools assigned to deliver PALS-UK will undertake training September 2019 in the schools region.
- Year 5 pupils will sit a reading attainment test prior to intervention in September 2019 (2nd - 14th) that will be administered by their classroom teacher and take 40 minutes to complete.
- Surveys will be sent out to all headteacher in all schools.
- The evaluation team will observe some training sessions.

At the end of school year (2019/20):

- A sub-sample of Year 5 pupils in all schools will undertake reading attainment and fluency test which will take 30 minutes.
- Year 5 pupils will answer a self-efficacy survey and reading attainment test.
- Surveys will be sent out to all head teachers and all Year 5 teachers
- Interviews with head teacher and Year 5 teachers will be conducted in several schools delivering PALS.

Milestone	Date
Schools offer Y5 parents opportunity to opt out from research	June/July 2019
Delivery team requests pupil data from schools	July 2019
Year 5 children take reading assessment tests	2nd to the 13th September 2019
Head teachers and teachers complete surveys	Sept/Oct 2019
Schools notified of allocation to intervention or control group	Sept/Oct 2019
Teachers from intervention schools attend training session	October 2019
Observation of training	Oct/Dec 2019

Teachers from the intervention schools participate at top up training	January 2021
Year 5 children take reading assessment, reading fluency and comprehension measure and complete self-efficacy survey	June 2020
Head teachers and Y5 teachers complete online surveys	June 2020
Small number of Head teacher and/or Y5 teachers participate in phone interviews	June 2020
Publication of evaluation report	December 2021

Head Teacher agreement

I agree for my school_____totake part in the **Evaluation of Peer Assisted Learning Strategies for Reading (PALS-UK)** and I accept the terms and conditions outlined in the Memorandum of Understanding and Privacy Notice.

School Name:

_____

Head Teacher Name:

_____

Head Teacher Signature: _____

Date:_____/____/_____

Head Teacher Email Address: _____

_____

Classroom Teacher Agreement

PALS-UK Lead Teacher Name:

_____

PALS-UK Lead Teacher Email Address: _____

Classroom Teacher Signature: _____

Date:_____/____/_____

PALS-UK Lead Teacher Job Title:_____

Contact details

School Contact (if not Head Teacher or Project

Lead):_____

School Contact Email Address (if not Head Teacher or interventionfacilitator):___

School Telephone Number:

_____

Thank you for agreeing to take part in this research project.

Please return this form to:

Completion of the MOU can be done either electronically or via post. Once completed either sign, scan and return the form electronically to Dr Emma Vardy at emma.vardy@coventry.ac.uk. Or if youwish to post the completed form return it to Dr Emma Vardy, Faculty of Health and Life Sciences, Richard Crossman Building 4th Floor, Coventry University, Priory Street, Coventry, CV1 5FB.

Original participant Information sheet and withdrawal form for parents

Evaluation of Peer Assisted Learning Strategies (PALS-UK)

PARTICIPANT INFORMATION SHEET FOR PARENTS/CARERS

The school your child attends has agreed to take part in a project assessing the effectiveness of a programme called Peer Assisted Learning Strategies (PALS-UK). This programme is designed to improve children's reading comprehension and reading fluency. The Head Teacher has agreed that all Year 5 will take part. Your child will be in Year 5 next year (Academic Year 2019/2020). We want to inform you about the project, and what it involves for you and your child.

We would like to collect data from your child's school records (such as name, gender, date of birth, pupil number, grades, test scores, etc.) and administer some reading tests. We will keep information confidential and only use it for research and evaluation purposes. Information about your child will be combined with that from other students to prepare reports. At no time will your child's identity or individual data be reported. It is the school's decision to deliver the PALS-UK reading programme, but it is your decision whether we can include your child's data in the evaluation. Participation in the evaluation is voluntary. Your decision will not adversely affect your child's grades or access to services or programs at your child's school. You are free to decline or stop at any time.

If you agree to let us include your child's data, you do NOT need to do anything else.
If you DO NOT want your child's data to contribute to this research, please complete the attached Withdrawal Form and return it to your child's teacher no later than 15th of July, 2019You can also complete the online withdrawal form here: [link to withdrawal form]

Before you make a decision, please read this Parent Information Sheet and the attached Privacy Notice carefully. Keep both sheets for future reference. If you have any questions, please speak with a member of the study team (contact information is on p. 4)

Who is undertaking the trial?

The trial is funded by the Education Endowment Foundation (EEF). The PALS-UK programme will be delivered by your child's usual classroom teacher, who will be supported by Dr Emma Vardy and Dr Helen Breadmore (Coventry University - the Delivery Team). The effectiveness of the intervention will be independently evaluated by RAND Europe (the Evaluation Team).

What is the purpose of the evaluation project?

Reading comprehension and reading fluency are foundational skills for learning. It is important to ensure that all children have the skills they need to support their learning in school. PALS-UK is a 20-week whole-class reading programme that aims to help children develop these skills. Existing research suggests that PALS-UK can have many benefits for pupils. For example, raising the profile of language in school, providing professional support for teaching and learning, and raising academic and social outcomes. This project aims to study the effectiveness of PALS-UK in Year 5.

We are conducting this study across 80 schools in England, as a Randomised Controlled Trial. This means half of the schools will be randomly allocated to take part in the PALS-UK programme. These are the ***'intervention'*** schools. The other half will not take part in PALS-UK. These are the ***'control'*** schools. Random allocation is the most robust way to evaluate how well the programme works. At

this stage we do not know whether will be an intervention or a control school. We will collect data from all schools, whether intervention or control.

Intervention schools: Will deliver PALS-UK in the academic year 2019/2020. Year 5 teachers receive the PALS-UK manual, training, resources and support needed to deliver the intervention. Year 5 teachers pair pupils and train them to use specific prompts, corrections and feedback during reading. Pupils participate in PALS-UK sessions three times a week, 30-35 minutes per session for 20 weeks. These sessions replace guided reading or similar reading activities.

Control schools: Teachers will not receive any coaching or training on PALS-UK and will teach literacy as "business as usual".

What is involved if my child takes part?

Summer- Autumn 2019: Recruitment and set-up (all schools)

- Schools share basic data on all children who will be in Year 5 in September 2019 with Delivery and Evaluation teams (see Privacy Notice for details).
- Year 5 pupils complete a reading attainment test (PIRA) and a reading self-efficacy survey in September 2019. This is administered by their classroom teacher and takes 40 minutes. We call this the pre-test. After this, schools are allocated to the intervention or control group.

Autumn – Spring 2019/2020:

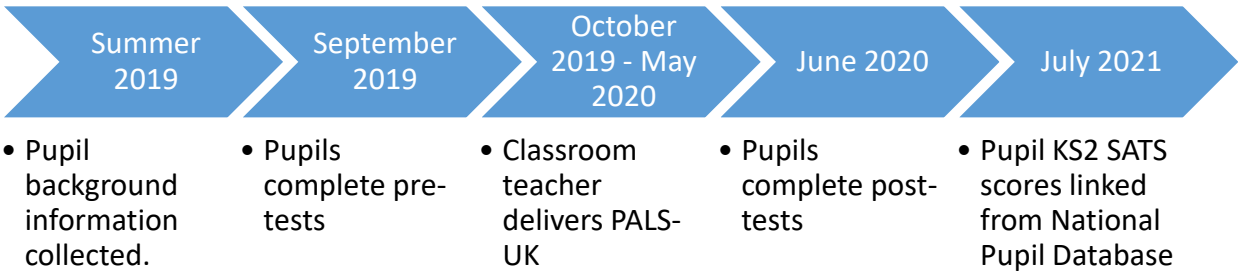
- Year 5 teachers in intervention schools receive PALS-UK training. Teachers deliver PALS-UK sessions three times a week for 30-35 minutes over 20 weeks (October 2019 - March 2020).

Summer 2020:

- In June 2020, all Year 5 pupils complete a reading assessment test (PIRA) and a self-efficacy survey. This takes about 40 minutes. Around one third of pupils also complete a reading comprehension and oral reading fluency assessment (WIAT-III subtests). This takes about 25 minutes. Assessors will be appropriately trained and have Enhanced Disclosure and Barring Service clearance (DBS). We call this the post-test.

Summer 2021:

- To study longer term impact, RAND Europe may use the National Pupil Database (held by the Department for Education) to collect your child's Key Stage 2 results in English.



What will happen to the results?

Once you child's information is included, the data will be anonymised. It will not be possible to identify individual pupils and their test results. We will not use your child's name or the name of the school in any reports. After the trial is complete, data will be archived. Follow-up studies may link pupil data from the National Pupil Database data using Pupil Matching References. All personal data will be treated with strictest confidence in accordance with the requirements of the General Data Protection Regulation (GDPR) and Data Protection Act 2018.

All data will be processed and stored securely on servers within the Europe Economic Area. At the end of the trial, data will be shared with the EEF archive, who will act as the data controller. Data will be shared with the Department for Education (the EEF's archive manager), the Office for National Statistics and potentially other research teams. Please see the Privacy Notice for detailed information. Results will be reported on the EEF website and will be freely available, and will also be published in academic journals and presented at conferences.

What are the risks and benefits of taking part?

Before contacting you we gained full support from the head teacher and classroom teacher of your school. We encouraged them to consider the risks and benefits. PALS-UK is designed to enhance normal classroom literacy activities. It is expected to have a positive impact on reading attainment.

Who has approved this research? What if I have a query or complaint?

This project has been peer reviewed and received ethical approval from the Faculty of Health and Life Sciences Research Ethics Committee of Coventry University and RAND Human Research Protection Program. It is conducted in accordance with British Psychological Society guidelines.

If you have any questions about how we will use your child's data or any concerns, please contact:

Dr Sashka Dimova at RAND Europe at **PALS@rand.org**, 01223353329

or

Dr Helen Breadmore at Coventry University **helen.breadmore@coventry.ac.uk** 024 7765 9512

If you have questions about your rights as a research participant or need to report a research-related injury or concern, you can contact RAND's Human Subjects Protection Committee toll-free at (868) 697-5620 or by emailing hspcinfo@rand.org. Please reference Study #XXXX-XXXX. You can also report concerns to Coventry University Ethics Committee by emailing **ethics.hls@coventry.ac.uk, please provide the reference** P90795.

Thank you very much for your time!

EVALUATION OF PEER ASSISTED LEARNING STRATEGIES FOR READING (PALS-UK)

RESEARCH PROJECT WITHDRAWAL SLIP

Please only complete if you **do NOT** want your child's data to be used in this research project.

If you are happy for your child's data to contribute to this project, then you do not have to do anything.

If you **do NOT** want your child's data to be used for this evaluation, you can withdraw your child via returning this form to the school or completing a form online [link to withdrawal form].

I, the undersigned, hereby **do NOT** give permission for my child's data to be used for the purposes of the evaluation of the Peer Assisted Learning Strategies (PALS-UK).

Child's full name: _____

School: _____ Class/form_____

Parent/guardian name: _____

Parent/guardian signature: _____

Date: _____

COVID-19 updated MoU

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

MEMORANDUM OF UNDERSTANDING

Addendum June 2020

This information is for schools taking part in the EEF funded Evaluation of Peer Assisted Learning Strategies (PALS-UK) (**https://vkc.mc.vanderbilt.edu/frg/about-3/**) research project in the 2019/20 and 2020/21 academic years. Please read this information carefully.

In light of the COVID-19 crisis, the data collection associated with the post-testing which was due to take place in May 2020 is now scheduled for the Autumn term 2020

Aims of the Evaluation

The aim of this project is to evaluate the effects of PALS-UK on reading comprehension and fluency for pupils in Year 5. Existing research suggests that PALS can benefit schools in multiple ways, such as raising the profile of language in primary schools and providing schools with professional support for teaching and learning. Furthermore, studies have found positive effects of peer tutoring in relation to reading skills, such as accuracy, comprehension and fluency. By participating in this evaluation, schools and individual participants contribute to furthering our knowledge on strategies for improving pupil's reading fluency and comprehension.

The evaluation is funded by the Education Endowment Foundation (EEF). The intervention PALS-UK is being delivered by Nottingham Trent University and Coventry University (the Delivery team) and will be independently evaluated by RAND Europe (the Evaluation team)[39] .

Changes due to COVID-19: Postponement of outcome measures

Due to COVID-19, intervention outcomes will be assessed by examining pupils' outcomes in the autumn term of 2020, instead of at the end of school year 2019/20. There are two changes to the MOU previously signed

1. Delay in the post-test timing. As planned, we will measure all pupils' overall reading and then a sub-sample of pupils' reading fluency, reading comprehension skills and reading self-efficacy. The data collectors will be appropriately trained and have full Disclosure and Barring Service clearance.
2. The independent evaluators will e-mail an additional survey in the autumn 2020 to headteachers and teachers. This survey will be the same as the survey completed in June 2020.

[39] **http://www.rand.org/randeurope.html**;

Head Teacher agreement

I agree for my school _____to continue to take part in the Evaluation of Peer Assisted Learning Strategies for Reading (PALS-UK) and I accept the terms and conditions outlined in the Memorandum of Understanding, Privacy Notice and Addendum.

School Name: _____

Head Teacher Name: _____

Head Teacher Signature: _____

Date: ____/____/_____

Head Teacher Email Address: _____

Classroom Teacher Agreement

PALS-UK Lead Teacher Name: _____

PALS-UK Lead Teacher Email Address: _____

Classroom Teacher Signature: _____

Date: ____/____/_____

PALS-UK Lead Teacher Job Title:_____

Contact details

School Contact (if not Head Teacher or Project

Lead):_____

School Contact Email Address (if not Head Teacher or intervention facilitator):_____

School Telephone Number: _____

Thank you for agreeing to take part in this research project.

Please return this form to:

Completion of the Addendum can be done either electronically or via post. Once completed either sign, scan (or digitally photograph) and return the form electronically to Dr Emma Vardy at emma.vardy@ntu.ac.uk. Or if you wish to post the completed form return it to Dr Emma Vardy, Nottingham Trent University, Psychology Department, School of Social Sciences, Room Chaucer 434,

50 Shakespeare Street, Nottingham, NG1 4FQ.

COVID-19 updated participant information sheet and withdrawal form

Evaluation of Peer Assisted Learning Strategies (PALS-UK)

PARTICIPANT INFORMATION SHEET FOR PARENTS/CARERS

Version 2. Updated June 2020 in light of the COVID-19 crisis

Updates have been highlighted in yellow.

Dear Parent/Guardian,

The school your child attends agreed to take part in a project assessing the effectiveness of a programme called Peer Assisted Learning Strategies (PALS-UK) in the academic year of 2019/20. This programme is designed to improve children's reading comprehension and reading fluency. The Head Teacher agreed that all Year 5 would take part, and teachers delivered the programme for 20 weeks. The post-testing for the project was due to take place in May 2020. However, this did not occur due to the COVID-19 crisis, so we want to inform you about the changes to the project, and what it involves for you and your child.

Unless you previously withdrew from the research project, we have already collected data from your child's school records (such as name, gender, date of birth, pupil number, grades, test scores, etc.) and administered some reading tests in September 2019. We are going to administer the same tests as planned, but this will be delayed until the autumn term of 2020 due to social distancing guidelines. Your child's school has asked we share with them the results of the autumn testing to provide a school baseline. We will share an overview of the data collected and if possible individual children's data to help schools in preparing to support your child's and their classmates learning for the new school year. The altered timeline is the main change from our original protocol. The additional change is sharing your child's data securely with their school.

Information about your child will be combined with that from other students to prepare reports. At no time will your child's identity or individual data be reported. It is your decision whether we can include your child's data in the post-test evaluation. Participation in the evaluation is voluntary. Your decision will not adversely affect your child's grades or access to services or programmes at your child's school. You or your child are free to decline or stop at any time.

If you agree to let us include your child's data, you do NOT need to do anything else.

We would like to remind you that you retain the right to withdraw if you DO NOT want your child to participate in the post-testing. To do so, please complete the attached Withdrawal Form and return it to your child's teacher no later than 30th of September 2020. You can also complete the online withdrawal form here: https://www.smartsurvey.co.uk/s/1608K/

Before you make a decision, please read the updated Parent Information Sheet and the updated Privacy Notice carefully. Keep both sheets for future reference. If you have any questions, please speak with a member of the study team (contact information is on p. 3). **You do not need to withdraw your child again if you have already done so.**

Who is undertaking the trial?

The trial is funded by the Education Endowment Foundation (EEF). The PALS-UK programme will be delivered by your child's usual classroom teacher, who will be supported by Dr Emma Vardy and Dr Helen Breadmore (Nottingham Trent University and Coventry University - the Delivery Team). The effectiveness of the intervention will be independently evaluated by RAND Europe (the Evaluation Team).

What is the purpose of the evaluation project?

Reading comprehension and reading fluency are foundational skills for learning. It is important to ensure that all children have the skills they need to support their learning in school. PALS-UK is a 20-week whole-class reading programme that aims to help children develop these skills. Existing research suggests that PALS-UK can have many benefits for

pupils. For example, raising the profile of language in school, providing professional support for teaching and learning, and raising academic and social outcomes. This project aims to study the effectiveness of PALS-UK in Year 5.

We are conducting this study across 89 schools in England, as a Randomised Controlled Trial. This means half of the schools have been randomly allocated to take part in the PALS-UK programme. These are the ***'intervention'*** schools. The other half is not taking part in PALS-UK. These are the ***'control'*** schools. Random allocation is the most robust way to evaluate how well the programme works.

Intervention schools: PALS-UK has been delivered in the academic year 2019/2020. Year 5 teachers received the PALS-UK manual, training, resources and support needed to deliver the intervention. Year 5 teachers paired pupils and trained them to use specific prompts, corrections and feedback during reading. Pupils participated in PALS-UK sessions three times a week, 30-35 minutes per session for 20 weeks. These sessions replace guided reading or similar reading activities.
Control schools: Teachers did not receive any coaching or training on PALS-UK and were teaching literacy as "business as usual".

What is involved if my child takes part?
Summer- Autumn 2019: Recruitment and set-up (all schools)

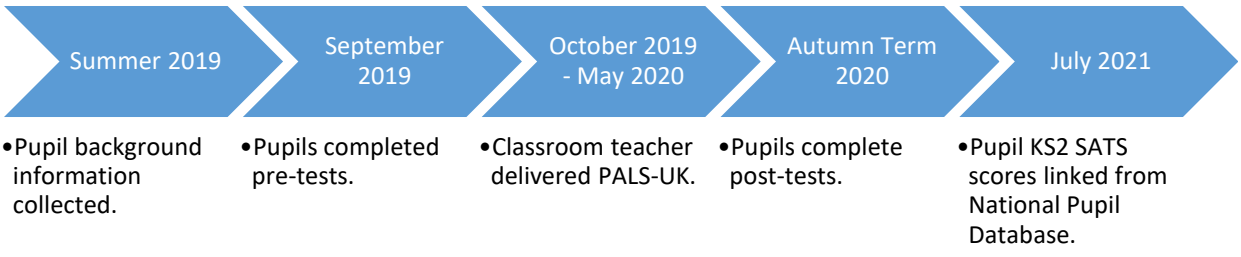
- Schools shared basic data on all children who were in Year 5 in September 2019 with the Delivery and Evaluation teams (see attached Privacy Notice for details).
- Year 5 pupils completed a reading attainment test (PIRA) and a reading self-efficacy survey in September 2019. This was administered by their classroom teacher and took 40 minutes. We are calling this the pre-test. After this, schools were allocated to the intervention or control group.

Autumn – Spring 2019/2020:

- Year 5 teachers in the intervention schools received PALS-UK training. Teachers deliver PALS-UK sessions three times a week for 30-35 minutes over 20 weeks (October 2019 - March 2020).

Autumn 2020 (delayed from Summer 2020 due to the COVID-19 crisis):

- In the autumn term of 2020, all Year 5 pupils will complete a reading assessment test (PIRA) and a self-efficacy survey. This takes about 40 minutes. Around one third of the pupils will also complete a reading comprehension and oral reading fluency assessment (WIAT-III subtests). This takes about 25 minutes. Assessors will be appropriately trained and have Enhanced Disclosure and Barring Service clearance (DBS). We call this the post-test.



What will happen to the results?
The pupil data will be anonymised once it is linked the National Pupil Database data using Pupil Matching References. It will not be possible to identify individual pupils and their test results. We will not use your child's name or the name of the school in any reports. After the trial is complete, data will be archived. Follow-up studies may link pupil data from the National Pupil Database data using Pupil Matching References. All personal data will be treated with strictest confidence in accordance with the requirements of the General Data Protection Regulation (GDPR) and Data Protection Act 2018.

All data will be processed and stored securely on servers within the Europe Economic Area. At the end of the trial, data will be shared with the EEF archive, who will act as the data controller. Data will be shared with the Department for Education, the EEF's archive manager, the Office for National Statistics and potentially other research teams. Please see the Privacy Notice for detailed information. Results will be reported on the EEF website and will be freely available, and will also be published in academic journals and presented at conferences.

What are the risks and benefits of taking part?

Before contacting you we gained full support from the head teacher and classroom teacher of your school. We encouraged them to consider the risks and benefits. PALS-UK is designed to enhance normal classroom literacy activities. It is expected to have a positive impact on reading attainment.

Who has approved this research? What if I have a query or complaint?
This project has been peer reviewed and received ethical approval from the Faculty of Health and Life Sciences Research Ethics Committee of Coventry University and RAND Human Research Protection Program. It is conducted in accordance with British Psychological Society guidelines.

If you have any questions about how we will use your child's data or any concerns, please contact:

Dr Sashka Dimova at RAND Europe at **PALS@randeurope.org**, 01223353329
or
Dr Helen Breadmore at Coventry University **helen.breadmore@coventry.ac.uk** 024 7765 9512

If you have questions about your rights as a research participant or need to report a research-related injury or concern, you can contact RAND's Human Subjects Protection Committee toll-free at (868) 697-5620 or by emailing hspcinfo@rand.org. Please reference Study IRB00000051. You can also report concerns to Coventry University Ethics Committee by emailing **ethics.hls@coventry.ac.uk, please provide the reference** P90795.

Thank you very much for your time!

EVALUATION OF PEER ASSISTED LEARNING STRATEGIES FOR READING (PALS-UK)

RESEARCH PROJECT WITHDRAWAL SLIP

Please only complete if you **DO NOT** want your child's data to be used in this research project.

If you are happy for your child's data to contribute to this project, then you do not have to do anything.

If you **DO NOT** want your child's data to be used for this evaluation, you can withdraw your child via returning this form to the school or completing a form online **https://www.smartsurvey.co.uk/s/1608K/**

I, the undersigned, hereby **DO NOT** give permission for my child's data to be used for the purposes of the evaluation of the Peer Assisted Learning Strategies (PALS-UK).

Child's full name: _____

School: _____ Class/form_____

Parent/guardian name: _____

Parent/guardian signature: _____

Date: _____

Appendix D: Baseline equivalence

School level

Figure 26 Distribution of school-level average KS2 Reading scores across the intervention and control groups among the sample as randomised

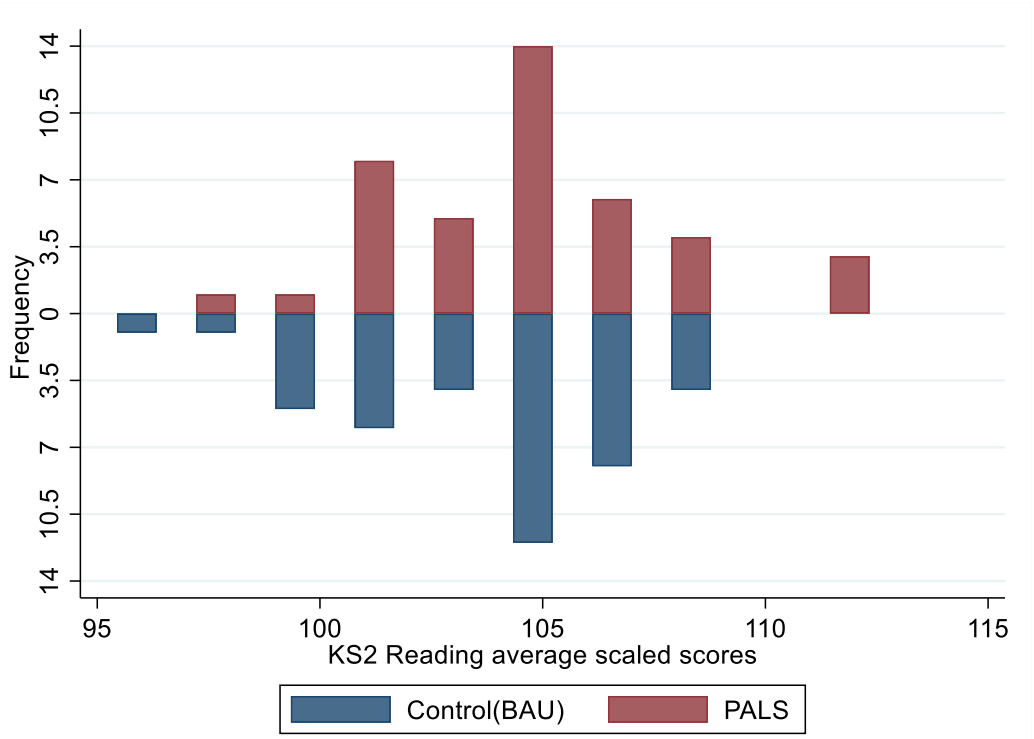
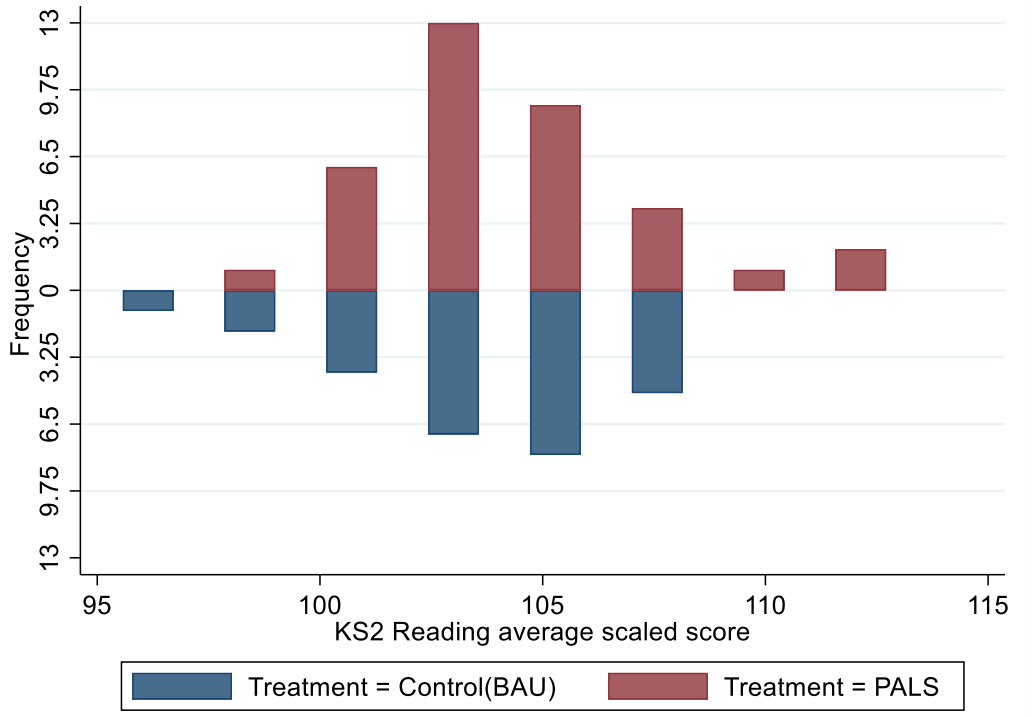


Figure 27 Distribution of school-level average KS2 Reading scores across the intervention and control groups among the sample as analysed



Pupil level

Figure 28 Distribution of age in months by treatment condition at baseline among the sample as randomised

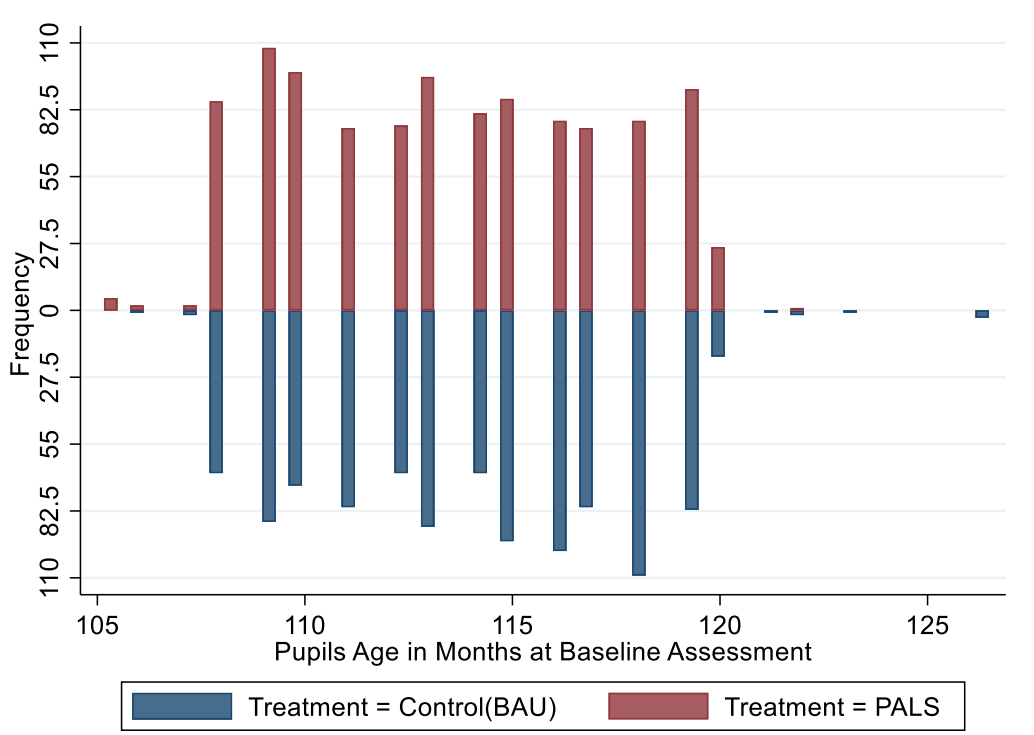


Figure 29 Distribution of age in months by treatment condition at baseline among the sample as analysed

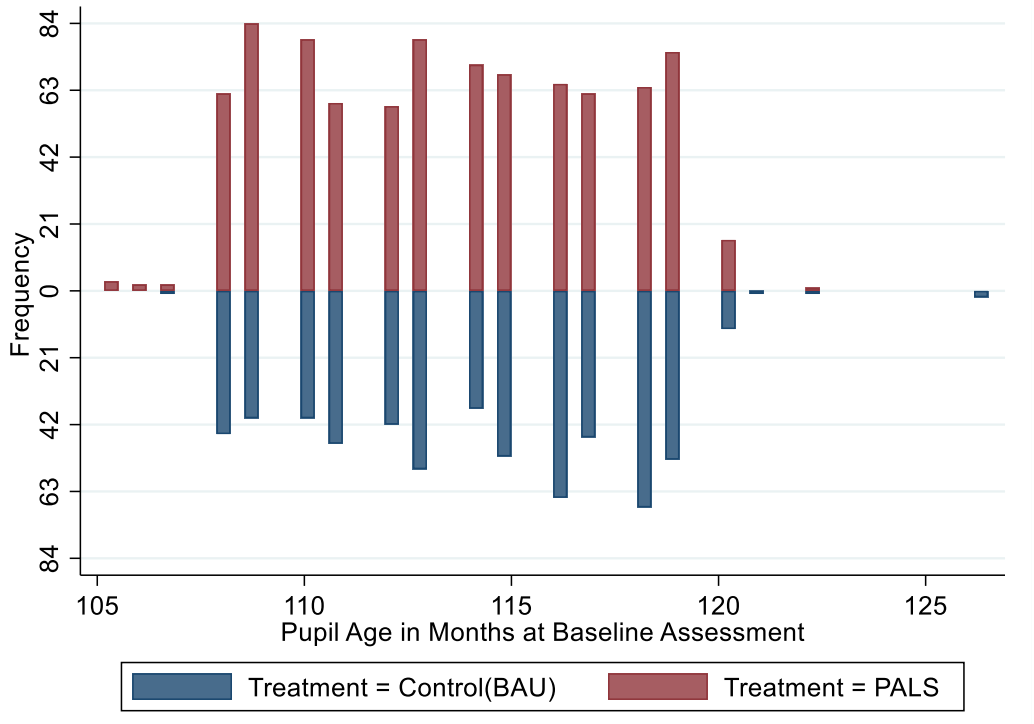


Figure 30 Distribution of age standardised PiRA scores by treatment condition at baseline among the sample as randomised

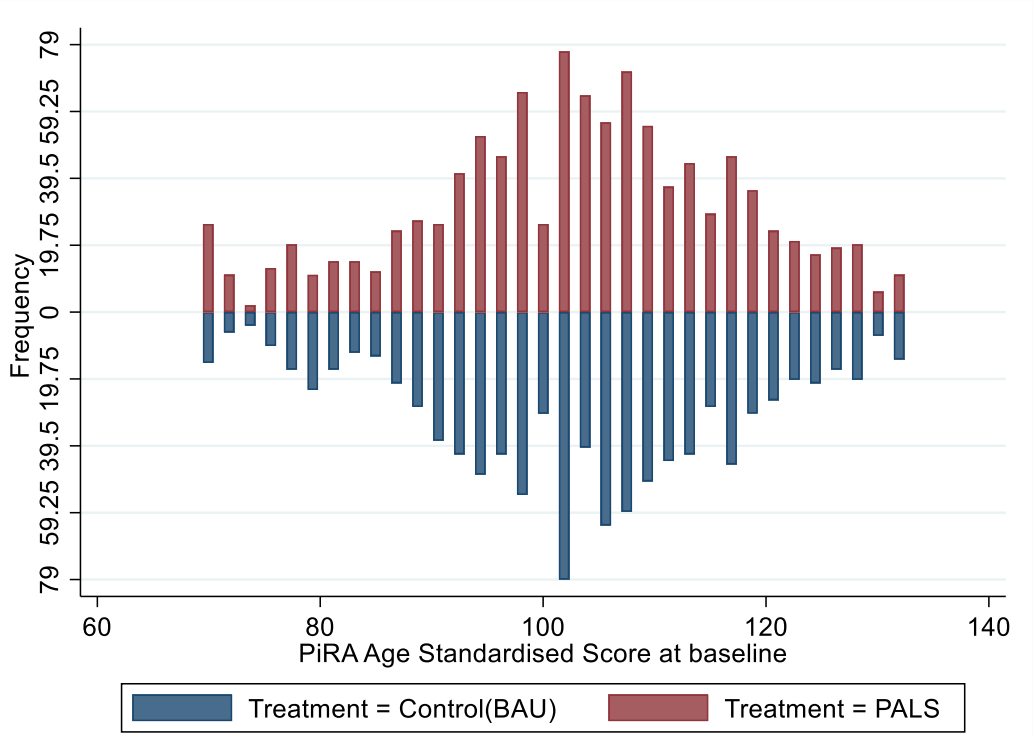


Figure 31 Distribution of age standardised PiRA scores by treatment condition at baseline among the sample as analysed

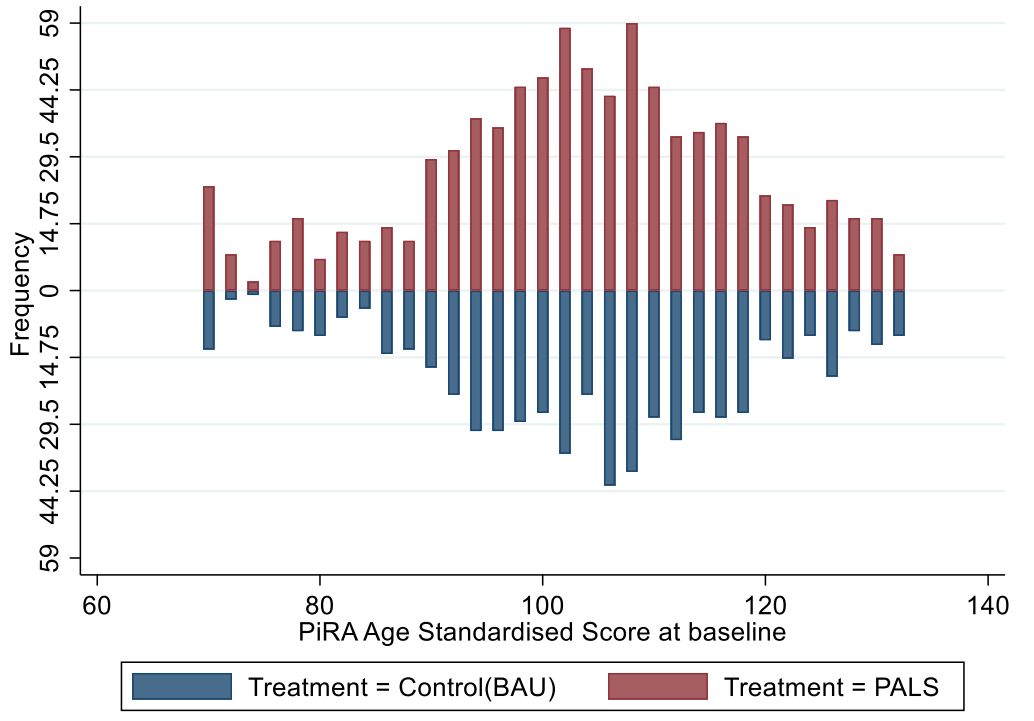


Figure 32 Distribution of total self-efficacy scores by treatment condition at baseline among the sample as randomised

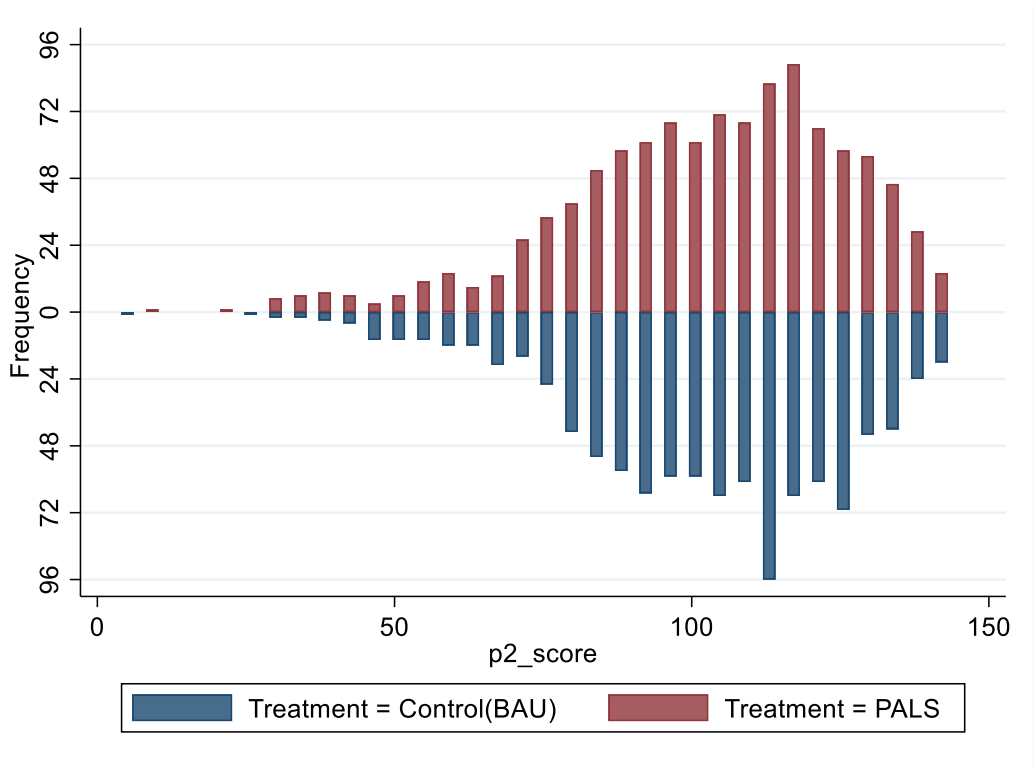
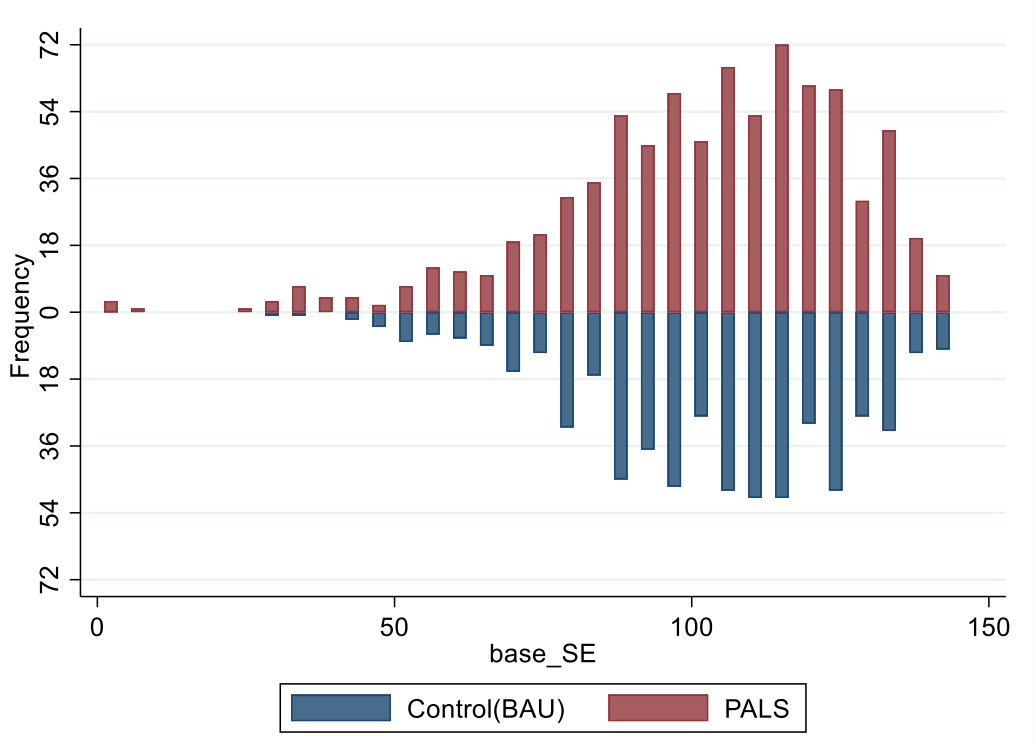
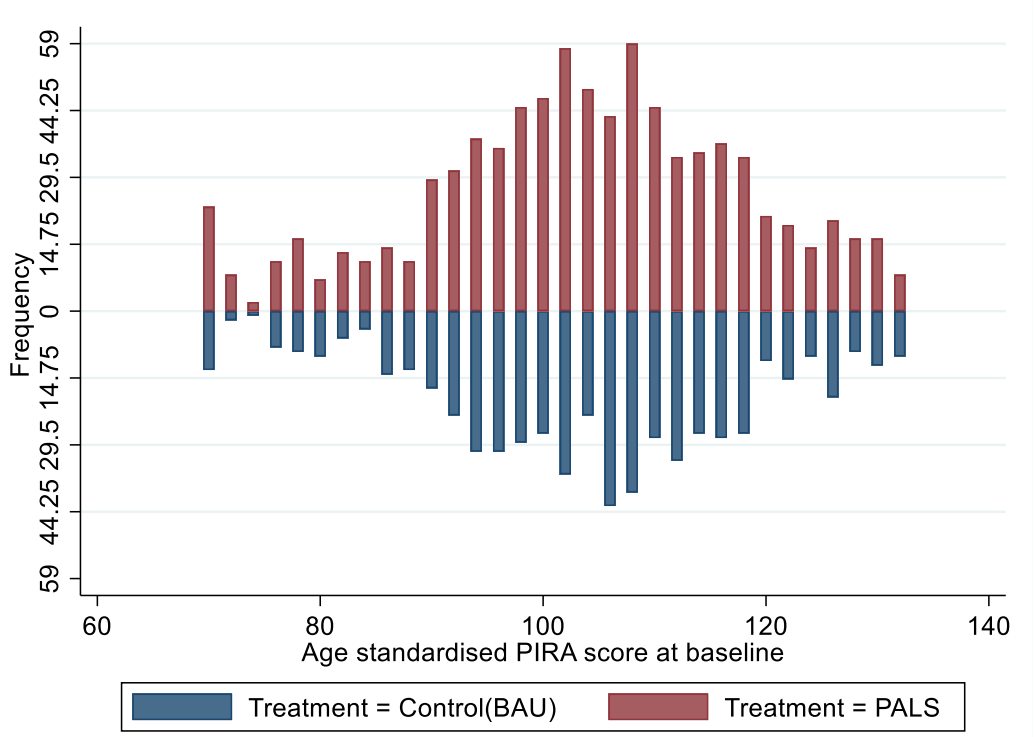


Figure 33 Distribution of total self-efficacy scores by treatment condition at baseline among the sample as analysed



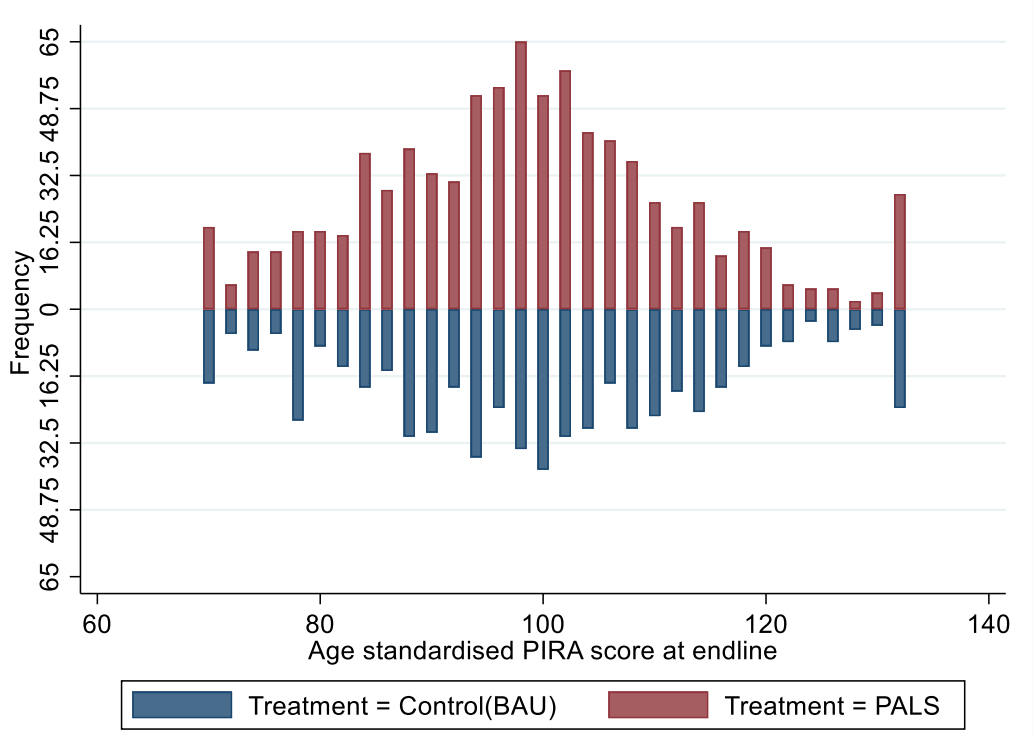
Appendix E: Distribution of PiRA scores in the final sample by treatment

Figure 34 Distribution of age standardised PiRA scores at baseline



Note: this graph includes only the N=1,459 pupils that were included in the final analytical model.

Figure 35 Distribution of age standardised PiRA scores at endline



Note: this graph includes only the N=1,459 pupils that were included in the final analytical model.

Appendix F: PALS-UK Reading self-efficacy questionnaire

Feelings about Reading Questionnaire

Part 1

In this part of the questionnaire we would like to know your thoughts on reading. There are no right or wrong answers; we just want to know your opinions. Please circle the answer that is most like you.

I enjoy reading	Really true for me	A bit true for me	A bit untrue for me	Really untrue for me
I only read to get rewards (for example stickers)	Really true for me	A bit true for me	A bit untrue for me	Really untrue for me
I like talking about the things I have read	Really true for me	A bit true for me	A bit untrue for me	Really untrue for me
I have lots of reading material at home	Really true for me	A bit true for me	A bit untrue for me	Really untrue for me
I spend time reading at home	Really true for me	A bit true for me	A bit untrue for me	Really untrue for me
I avoid reading if I can	Really true for me	A bit true for me	A bit untrue for me	Really untrue for me
I think reading is important	Really true for me	A bit true for me	A bit untrue for me	Really untrue for me
I get excited about reading a book that I want to read more	Really true for me	A bit true for me	A bit untrue for me	Really untrue for me
Knowing how to read well is important to me	Really true for me	A bit true for me	A bit untrue for me	Really untrue for me
I like it when it is time to read in class	Really true for me	A bit true for me	A bit untrue for me	Really untrue for me

Part 2: Practice

Read each item and rate how certain **YOU** are that you can do the actions described below by circling one of the numbers below the item. High scores equal a higher certainty that you can do the action. Please wait for instructions before starting the practice items.

Lift a bag of sugar

Very certain I cannot do 1 2 3 4 5 6 7 Very certain I can do

Lift one of my class friends

Very certain I cannot do 1 2 3 4 5 6 7 Very certain I can do

Lift two of my class friends

Very certain I cannot do 1 2 3 4 5 6 7 Very certain I can do

Main questionnaire

Read each sentence and rate how certain **YOU** are that you can do the things described below. It is important you tell us what **YOU** think about your reading. When you think about reading, think about the any reading that you do at school and at home. These could be things you read in books, magazines, newspapers, comics, emails, text messages and the internet. To give an answer circle one of the numbers on the scale below the item.

1. Read out loud in front of the class

Very certain I cannot do 1 2 3 4 5 6 7 Very certain I can do

2. Continue reading even when I find it difficult

Very certain I cannot do 1 2 3 4 5 6 7 Very certain I can do

3. Practice reading in my spare time even when I don't have to

Very certain I cannot do 1 2 3 4 5 6 7 Very certain I can do

4. Sound out a word that I find hard to read

Very certain I cannot do 1 2 3 4 5 6 7 Very certain I can do

5. Read on my own without an adult's help

Very certain I cannot do 1 2 3 4 5 6 7 Very certain I can do

6. Read things that are harder than the book I normally read at school

Very certain I cannot do 1 2 3 4 5 6 7 Very certain I can do

7. Know what I need to do to improve my reading

Very certain I cannot do 1 2 3 4 5 6 7 Very certain I can do

8. Continue reading even when I find it boring (e.g., a school book or magazine)

Very certain I cannot do 1 2 3 4 5 6 7 Very certain I can do

9. Read out loud quickly and still get words right

Very certain I cannot do 1 2 3 4 5 6 7 Very certain I can do

10. Improve my reading if I want to

Very certain I cannot do 1 2 3 4 5 6 7 Very certain I can do

11. Continue reading even when I do not like the subject

Very certain I cannot do 1 2 3 4 5 6 7 Very certain I can do

12. Read as well as my friends

Very certain I cannot do 1 2 3 4 5 6 7 Very certain I can do

13. Continue reading even when I get frustrated

Very certain I cannot do 1 2 3 4 5 6 7 Very certain I can do

14. Work out the sounds in words I have not seen before

Very certain I cannot do 1 2 3 4 5 6 7 Very certain I can do

15. Read without making lots of mistakes

Very certain I cannot do 1 2 3 4 5 6 7 Very certain I can do

16. Understand difficult books

Very certain I cannot do 1 2 3 4 5 6 7 Very certain I can do

17. Read a book I have not read before

Very certain I cannot do 1 2 3 4 5 6 7 Very certain I can do

18. Make good predictions of what will happen next in a story

Very certain I cannot do 1 2 3 4 5 6 7 Very certain I can do

19. Summarise a book I have read

Very certain I cannot do 1 2 3 4 5 6 7 Very certain I can do

20. Read a book that I think is a hard, challenging book

Very certain I cannot do 1 2 3 4 5 6 7 Very certain I can do

Appendix G: Randomisation code, analysis code and output

Primary analysis

```
. xtmixed agesscore_PIRA_end ib0.Treatment agesscore_PIRA_base i.area_form_strata || sch_harmonised_num: if agesscore_PIRA_base!=.
>  & agesscore_PIRA_end!=.

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log likelihood = -5527.8754
Iteration 1:   log likelihood = -5527.8754

Computing standard errors:

Mixed-effects ML regression                     Number of obs     =      1,459
Group variable: sch_harmonis~m                  Number of groups  =         66
                                                Obs per group:
                                                              min =          5
                                                              avg =       22.1
                                                              max =         51
                                                Wald chi2(5)      =     994.49
Log likelihood = -5527.8754                     Prob > chi2       =     0.0000
```

agesscore_PIRA_end	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
Treatment						
PALS	-.8338776	1.174664	-0.71	0.478	-3.136176	1.468421
agesscore_PIRA_base	.6467873	.0206163	31.37	0.000	.6063802	.6871945
area_form_strata						
Midlands/>1 class	-.8612394	1.394311	-0.62	0.537	-3.594038	1.871559
North East/1 class	-1.161426	1.565666	-0.74	0.458	-4.230075	1.907224
North East/>1 class	-.0466527	2.30419	-0.02	0.984	-4.562782	4.469476
_cons	32.56831	2.485236	13.10	0.000	27.69734	37.43928

Random-effects parameters	Estimate	Std. err.	[95% conf.	interval]
sch_harmon~m: Identity				
sd(_cons)	4.049759	.4543693	3.250329	5.045813
sd(Residual)	10.35335	.1958665	9.976492	10.74445

LR test vs. linear model: chibar2(01) = 129.33 Prob >= chibar2 = 0.0000

Secondary analysis

Reading self-efficacy

```
. xtmixed sum_endline1 treatment sum_baseline i.strata || school_num:, mle var

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log likelihood = -6150.3093
Iteration 1:   log likelihood = -6150.3093

Computing standard errors:

Mixed-effects ML regression                     Number of obs     =      1,372
Group variable: school_num                      Number of groups  =         62

                                                Obs per group:
                                                              min =          6
                                                              avg =       22.1
                                                              max =         53

                                                Wald chi2(5)      =     242.73
Log likelihood = -6150.3093                     Prob > chi2       =     0.0000
```

sum_endline1	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
treatment	-.0514792	2.50292	-0.02	0.984	-4.957112	4.854153
sum_baseline	.3520463	.0227349	15.48	0.000	.3074866	.396606
strata						
MidlandsMidlands	-3.742723	2.985087	-1.25	0.210	-9.593386	2.107939
North East	-1.904234	3.278601	-0.58	0.561	-8.330173	4.521705
North EastNort..	-1.03495	5.244693	-0.20	0.844	-11.31436	9.24446
_cons	85.35132	3.940476	21.66	0.000	77.62813	93.07451

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
school_num: Identity				
var(_cons)	73.09772	17.21229	46.07568	115.9674
var(Residual)	427.784	16.72489	396.2283	461.8528

LR test vs. linear model: chibar2(01) = 111.91 Prob >= chibar2 = 0.0000

Analysis in the presence of non-compliance

```
. reg fidelity_SAP Treatment
```

Source	SS	df	MS	Number of obs	=	2,292
				F(1, 2290)	=	694.54
Model	84.442418	1	84.442418	Prob > F	=	0.0000
Residual	278.419711	2,290	.12158066	R-squared	=	0.2327
				Adj R-squared	=	0.2324
Total	362.862129	2,291	.158385914	Root MSE	=	.34868

fidelity_SAP	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Treatment	.3840272	.0145718	26.35	0.000	.3554518	.4126025
_cons	-8.13e-15	.0104423	-0.00	1.000	-.0204773	.0204773

```
. mixed agesscore_PIRA_end fidres_SAP agesscore_PIRA_base i.area_form_strata || sch_harmonised_num:, mle var

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log likelihood = -5527.9719
Iteration 1:   log likelihood = -5527.9719

Computing standard errors:

Mixed-effects ML regression                     Number of obs     =      1,459
Group variable: sch_harmonis~m                  Number of groups  =         66

                                                Obs per group:
                                                              min =          5
                                                              avg =       22.1
                                                              max =         51

                                                Wald chi2(5)      =     994.17
Log likelihood = -5527.9719                     Prob > chi2       =     0.0000
```

agesscore_PIRA_~d	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
fidres_SAP	.8931936	1.600905	0.56	0.577	-2.244523	4.03091
agesscore_PIRA_~e	.6467754	.0206178	31.37	0.000	.6063652	.6871857
area_form_strata						
Midlands/>1 cl..	-.9983228	1.384477	-0.72	0.471	-3.711849	1.715203
North East/1 c..	-1.403807	1.607233	-0.87	0.382	-4.553926	1.746311
North East/>1 ..	-.0425779	2.305227	-0.02	0.985	-4.560741	4.475585
_cons	32.18932	2.410378	13.35	0.000	27.46507	36.91357

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
sch_harmon~m: Identity				
var(_cons)	16.38187	3.690406	10.53441	25.47516
var(Residual)	107.2109	4.057109	99.54685	115.465

LR test vs. linear model: chibar2(01) = 125.25 Prob >= chibar2 = 0.0000

Additional analysis in the presence of compliance

Compliance score 2

```
.         reg fidelity1 Treatment
```

Source	SS	df	MS	Number of obs	=	2,292
				F(1, 2290)	=	1074.97
Model	138.082763	1	138.082763	Prob > F	=	0.0000
Residual	294.15633	2,290	.128452546	R-squared	=	0.3195
				Adj R-squared	=	0.3192
Total	432.239092	2,291	.188668308	Root MSE	=	.3584

fidelity1	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Treatment	.491079	.014978	32.79	0.000	.4617072	.5204508
_cons	-1.22e-15	.0107333	-0.00	1.000	-.021048	.021048

```
. mixed agesscore_PIRA_end fidres1 agesscore_PIRA_base i.area_form_strata || sch_harmonised_num:, mle var

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log likelihood = -5527.6557
Iteration 1:   log likelihood = -5527.6557

Computing standard errors:

Mixed-effects ML regression                     Number of obs     =      1,459
Group variable: sch_harmonis~m                  Number of groups  =         66

                                                Obs per group:
                                                              min =          5
                                                              avg =       22.1
                                                              max =         51

                                                Wald chi2(5)      =     995.26
Log likelihood = -5527.6557                     Prob > chi2       =     0.0000
```

agesscore_PIRA_~d	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
fidres1	1.514174	1.554679	0.97	0.330	-1.532941	4.561289
agesscore_PIRA_~e	.6474359	.0206226	31.39	0.000	.6070163	.6878555
area_form_strata						
Midlands/>1 cl..	-1.047075	1.381037	-0.76	0.448	-3.753857	1.659707
North East/1 c..	-1.400336	1.572962	-0.89	0.373	-4.483284	1.682612
North East/>1 ..	.3250622	2.308318	0.14	0.888	-4.199158	4.849282
_cons	32.09492	2.406661	13.34	0.000	27.37795	36.81188

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
sch_harmon~m: Identity				
var(_cons)	16.23313	3.655895	10.44002	25.24079
var(Residual)	107.1972	4.056155	99.53491	115.4493

LR test vs. linear model: chibar2(01) = 125.32 Prob >= chibar2 = 0.0000

Compliance score 3

```
.         reg fidelity3 Treatment
```

Source	SS	df	MS	Number of obs	=	2,292
				F(1, 2290)	=	3535.65
Model	331.077648	1	331.077648	Prob > F	=	0.0000
Residual	214.435004	2,290	.09363974	R-squared	=	0.6069
				Adj R-squared	=	0.6067
Total	545.512653	2,291	.238111154	Root MSE	=	.30601

fidelity3	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Treatment	.7604078	.0127883	59.46	0.000	.73533	.7854856
_cons	-8.10e-15	.0091642	-0.00	1.000	-.0179709	.0179709

```
.         reg fidelity3 Treatment
```

Source	SS	df	MS	Number of obs	=	2,292
				F(1, 2290)	=	1748.88
Model	213.669027	1	213.669027	Prob > F	=	0.0000
Residual	279.780799	2,290	.122175021	R-squared	=	0.4330
				Adj R-squared	=	0.4328
Total	493.449825	2,291	.215386218	Root MSE	=	.34954

fidelity3	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
Treatment	.6108751	.0146074	41.82	0.000	.58223	.6395202
_cons	-1.12e-14	.0104678	-0.00	1.000	-.0205273	.0205273

```
. mixed agesscore_PIRA_end fidres3 agesscore_PIRA_base i.area_form_strata || sch_harmonised_num:, mle var

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log likelihood = -5527.0178
Iteration 1:   log likelihood = -5527.0178

Computing standard errors:

Mixed-effects ML regression                     Number of obs     =      1,459
Group variable: sch_harmonis~m                  Number of groups  =         66

                                                Obs per group:
                                                              min =          5
                                                              avg =       22.1
                                                              max =         51

                                                Wald chi2(5)      =     997.51
Log likelihood = -5527.0178                     Prob > chi2       =     0.0000
```

agesscore_PIRA_~d	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
fidres3	-2.779304	1.849731	-1.50	0.133	-6.404711	.8461019
agesscore_PIRA_~e	.6466987	.0206049	31.39	0.000	.6063138	.6870835
area_form_strata						
Midlands/>1 cl..	-.8646646	1.36709	-0.63	0.527	-3.544112	1.814783
North East/1 c..	-1.103322	1.544938	-0.71	0.475	-4.131345	1.924702
North East/>1 ..	.3571618	2.277084	0.16	0.875	-4.105841	4.820165
_cons	32.10594	2.401429	13.37	0.000	27.39923	36.81266

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
sch_harmon~m: Identity				
var(_cons)	15.80692	3.578313	10.14274	24.63424
var(Residual)	107.2001	4.056024	99.538	115.4519

LR test vs. linear model: chibar2(01) = 124.37 Prob >= chibar2 = 0.0000

```
. mixed agesscore_PIRA_end fidres3 agesscore_PIRA_base i.area_form_strata || sch_harmonised_num:, mle var

Performing EM optimization ...

Performing gradient-based optimization:
Iteration 0:   log likelihood = -5527.5401
Iteration 1:   log likelihood = -5527.5401

Computing standard errors ...

Mixed-effects ML regression                     Number of obs     =      1,459
Group variable: sch_harmonis~m                  Number of groups  =         66
                                                Obs per group:
                                                              min =          5
                                                              avg =       22.1
                                                              max =         51
                                                Wald chi2(5)      =     995.67
Log likelihood = -5527.5401                     Prob > chi2       =     0.0000
```

agesscore_PIRA_end	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
fidres3	1.86419	1.712803	1.09	0.276	-1.492841	5.221221
agesscore_PIRA_base	.6473443	.0206172	31.40	0.000	.6069354	.6877533
area_form_strata						
Midlands/>1 class	-.6849992	1.401855	-0.49	0.625	-3.432585	2.062586
North East/1 class	-1.306298	1.559055	-0.84	0.402	-4.36199	1.749395
North East/>1 class	.708979	2.364597	0.30	0.764	-3.925546	5.343504
_cons	31.87834	2.41655	13.19	0.000	27.14199	36.61469

Random-effects parameters	Estimate	Std. err.	[95% conf.	interval]
sch_harmon~m: Identity				
var(_cons)	16.14177	3.641191	10.37387	25.11664
var(Residual)	107.2008	4.056335	99.53823	115.4533

LR test vs. linear model: chibar2(01) = 124.56 Prob >= chibar2 = 0.0000

Missing data analysis

Main missing data analysis model

```
. melogit binary_PIRA_missing ib0.Treatment agesscore_PIRA_base ib2.area_classes_strata EVERFSM_6_SPR19 EAL_binary
> _var sch_FSM_percent sch_EAL_percent || MOU_Number:, or

Fitting fixed-effects model:

Iteration 0:   log likelihood = -1275.6808
Iteration 1:   log likelihood = -1275.1352
Iteration 2:   log likelihood = -1275.1351

Refining starting values:

Grid node 0:   log likelihood = -744.85557

Fitting full model:

Iteration 0:   log likelihood = -744.85557
Iteration 1:   log likelihood = -663.79449
Iteration 2:   log likelihood = -643.88191
Iteration 3:   log likelihood = -637.71441
Iteration 4:   log likelihood = -636.68776
Iteration 5:   log likelihood = -636.75443
Iteration 6:   log likelihood = -636.81478
Iteration 7:   log likelihood = -636.84387
Iteration 8:   log likelihood = -636.86124
Iteration 9:   log likelihood = -636.87101
Iteration 10:  log likelihood = -636.87631
Iteration 11:  log likelihood = -636.87914
Iteration 12:  log likelihood = -636.88063
Iteration 13:  log likelihood = -636.88142
Iteration 14:  log likelihood = -636.88184
Iteration 15:  log likelihood = -636.88205
Iteration 16:  log likelihood = -636.88217
Iteration 17:  log likelihood = -636.88223

Mixed-effects logistic regression               Number of obs     =      2,163
Group variable:      MOU_Number                 Number of groups  =         89

                                                Obs per group:
                                                              min =          6
                                                              avg =       24.3
                                                              max =         64

Integration method: mvaghermite                 Integration pts.  =          7

                                                Wald chi2(9)      =      26.98
Log likelihood = -636.88223                     Prob > chi2       =     0.0014
------------------------------------------------------------------------------
 binary_PIRA_missing | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------------+--------------------------------------------------------
           Treatment |
                PALS |   .1475201   .1008692    -2.80   0.005     .0386215    .5634736
 agesscore_PIRA_base |   .9877005   .0064643    -1.89   0.059     .9751117    1.000452
                     |
 area_classes_strata |
   Midlands 1 class  |   .0943208   .0795947    -2.80   0.005     .0180427    .4930746
 North East 1 class  |   .1624549   .1585447    -1.86   0.063     .0239894    1.100136
North East >1 class  |   .1012864   .1429903    -1.62   0.105     .0063661    1.611493
                     |
     EVERFSM_6_SPR19 |   1.457949   .2889223     1.90   0.057      .988688    2.149934
      EAL_binary_var |   1.474347    .438072     1.31   0.191     .8235359     2.63947
     sch_FSM_percent |   1.003045   .0189201     0.16   0.872     .9666392    1.040821
     sch_EAL_percent |   1.002997   .0151026     0.20   0.842     .9738292    1.033039
               _cons |   13.81168   15.00415     2.42   0.016     1.642683    116.1286
---------------------+--------------------------------------------------------
MOU_Number           |
          var(_cons)|   15.58226   3.979881                      9.445487    25.70613
------------------------------------------------------------------------------
Note: Estimates are transformed only in the first equation.
Note: _cons estimates baseline odds (conditional on zero random effects).
LR test vs. logistic model: chibar2(01) = 1276.51     Prob >= chibar2 = 0.0000
```

Missing data analysis model with an interaction term for treatment allocation and FSM status

```
. melogit binary_PIRA_missing agesscore_PIRA_base ib2.area_classes_strata ib3.FSM_Treatment_Inter EAL_binary_var s
> ch_FSM_percent sch_EAL_percent || MOU_Number:, or

Fitting fixed-effects model:

Iteration 0:   log likelihood =  -1274.189
Iteration 1:   log likelihood = -1273.5594
Iteration 2:   log likelihood = -1273.5593

Refining starting values:

Grid node 0:   log likelihood =  -745.2453

Fitting full model:

Iteration 0:   log likelihood =  -745.2453
Iteration 1:   log likelihood = -669.31295
Iteration 2:   log likelihood = -644.71332
Iteration 3:   log likelihood = -637.58225
Iteration 4:   log likelihood =  -636.6748
Iteration 5:   log likelihood = -636.72685
Iteration 6:   log likelihood =  -636.7816
Iteration 7:   log likelihood = -636.80976
Iteration 8:   log likelihood = -636.82662
Iteration 9:   log likelihood = -636.83607
Iteration 10:  log likelihood = -636.84119
Iteration 11:  log likelihood = -636.84392
Iteration 12:  log likelihood = -636.84536
Iteration 13:  log likelihood = -636.84612
Iteration 14:  log likelihood = -636.84652
Iteration 15:  log likelihood = -636.84673
Iteration 16:  log likelihood = -636.84684
Iteration 17:  log likelihood =  -636.8469

Mixed-effects logistic regression               Number of obs     =      2,163
Group variable:      MOU_Number                 Number of groups  =         89

                                                Obs per group:
                                                              min =          6
                                                              avg =       24.3
                                                              max =         64

Integration method: mvaghermite                 Integration pts.  =          7

                                                Wald chi2(10)     =      27.03
Log likelihood =  -636.8469                     Prob > chi2       =     0.0026
------------------------------------------------------------------------------
  binary_PIRA_missing | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------------------+-------------------------------------------------------
  agesscore_PIRA_base |   .9877208   .0064665    -1.89   0.059     .9751278    1.000477
                      |
  area_classes_strata |
    Midlands 1 class  |   .0942475   .0795419    -2.80   0.005     .0180253    .4927848
  North East 1 class  |   .1620037   .1581292    -1.86   0.062     .0239156     1.09741
 North East >1 class  |   .1007528   .1422764    -1.63   0.104     .0063277    1.604231
                      |
  FSM_Treatment_Inter |
Non FSM Control(BAU)  |   .6486391   .1884825    -1.49   0.136     .3669947    1.146427
       Non FSM PALS   |   .0985062   .0711695    -3.21   0.001     .0239046     .405925
      FSM pupil PALS  |   .1371711   .1014907    -2.68   0.007     .0321715    .5848617
                      |
       EAL_binary_var |   1.483766   .4426934     1.32   0.186     .8268039    2.662736
      sch_FSM_percent |   1.003065   .0189209     0.16   0.871     .9666579    1.040844
      sch_EAL_percent |   1.002877   .0151062     0.19   0.849      .973702    1.032926
                _cons |   20.95588   23.03135     2.77   0.006     2.431092    180.6385
----------------------+-------------------------------------------------------
MOU_Number            |
           var(_cons)|    15.5866   3.982052                     9.446848    25.71672
------------------------------------------------------------------------------
Note: Estimates are transformed only in the first equation.
Note: _cons estimates baseline odds (conditional on zero random effects).
LR test vs. logistic model: chibar2(01) = 1273.42     Prob >= chibar2 = 0.0000
```

Missing data analysis model with an interaction term for treatment allocation and EAL status

```
. melogit binary_PIRA_missing agesscore_PIRA_base ib2.area_classes_strata EVERFSM_6_SPR19 ib3.EAL_Treatment_Inter
> sch_FSM_percent sch_EAL_percent || MOU_Number:, or

Fitting fixed-effects model:

Iteration 0:   log likelihood =  -1262.664
Iteration 1:   log likelihood = -1261.7551
Iteration 2:   log likelihood = -1261.7549
Iteration 3:   log likelihood = -1261.7549

Refining starting values:

Grid node 0:   log likelihood = -745.55159

Fitting full model:

Iteration 0:   log likelihood = -745.55159
Iteration 1:   log likelihood = -673.75208
Iteration 2:   log likelihood = -647.95478
Iteration 3:   log likelihood = -639.06697
Iteration 4:   log likelihood = -636.81031
Iteration 5:   log likelihood = -636.71298
Iteration 6:   log likelihood = -636.79301
Iteration 7:   log likelihood = -636.83175
Iteration 8:   log likelihood = -636.85412
Iteration 9:   log likelihood =  -636.8671
Iteration 10:  log likelihood = -636.87423
Iteration 11:  log likelihood = -636.87806
Iteration 12:  log likelihood =  -636.8801
Iteration 13:  log likelihood = -636.88117
Iteration 14:  log likelihood = -636.88173
Iteration 15:  log likelihood = -636.88203
Iteration 16:  log likelihood = -636.88218
Iteration 17:  log likelihood = -636.88227
Iteration 18:  log likelihood = -636.88231

Mixed-effects logistic regression               Number of obs     =      2,163
Group variable:      MOU_Number                 Number of groups  =         89

                                                Obs per group:
                                                              min =          6
                                                              avg =       24.3
                                                              max =         64

Integration method: mvaghermite                 Integration pts.  =          7

                                                Wald chi2(10)     =      26.98
Log likelihood = -636.88231                     Prob > chi2       =     0.0026
------------------------------------------------------------------------------
  binary_PIRA_missing | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------------------+-------------------------------------------------------
  agesscore_PIRA_base |   .9877009   .0064652    -1.89   0.059     .9751103    1.000454
                      |
  area_classes_strata |
    Midlands 1 class  |   .0943179   .0796017    -2.80   0.005     .0180386    .4931569
  North East 1 class  |   .1624278   .1585837    -1.86   0.063     .0239665    1.100821
 North East >1 class  |   .1012804   .1429854    -1.62   0.105     .0063653    1.611509
                      |
      EVERFSM_6_SPR19 |   1.457873   .2894219     1.90   0.058     .9879527     2.15131
                      |
  EAL_Treatment_Inter |
Non EAL Control(BAU)  |   .6791092   .2794913    -0.94   0.347     .3031257    1.521446
       Non EAL PALS   |   .1001351   .0772083    -2.98   0.003      .022094    .4538348
      EAL pupil PALS  |   .1478226   .1230907    -2.30   0.022     .0289036    .7560138
                      |
      sch_FSM_percent |   1.003048   .0189317     0.16   0.872     .9666208    1.040849
      sch_EAL_percent |   1.002996    .015103     0.20   0.843     .9738272    1.033038
                _cons |   20.34247    23.3066     2.63   0.009     2.153637    192.1476
----------------------+-------------------------------------------------------
MOU_Number            |
           var(_cons)|   15.58337   3.987852                      9.437032    25.73283
------------------------------------------------------------------------------
Note: Estimates are transformed only in the first equation.
Note: _cons estimates baseline odds (conditional on zero random effects).
LR test vs. logistic model: chibar2(01) = 1249.75     Prob >= chibar2 = 0.0000
```

Results from Little's MCAR test

```
. mcartest agesscore_PIRA_end Treatment agesscore_PIRA_base area_classes_strata if Treatment!=., emoutput

Iteration 0:   Observed log likelihood = -11940.691
Iteration 1:   Observed log likelihood = -11640.157
Iteration 2:   Observed log likelihood = -11576.046
Iteration 3:   Observed log likelihood = -11564.278
Iteration 4:   Observed log likelihood = -11562.498
Iteration 5:   Observed log likelihood = -11562.252
Iteration 6:   Observed log likelihood = -11562.218
Iteration 7:   Observed log likelihood = -11562.214
Iteration 8:   Observed log likelihood = -11562.213
Iteration 9:   Observed log likelihood = -11562.213
Iteration 10:  Observed log likelihood = -11562.213
Iteration 11:  Observed log likelihood = -11562.213
Iteration 12:  Observed log likelihood = -11562.213
Iteration 13:  Observed log likelihood = -11562.213
Iteration 14:  Observed log likelihood = -11562.213
Iteration 15:  Observed log likelihood = -11562.213

Expectation-maximization estimation      Number obs           =      2275
                                         Number missing       =       872
                                         Number patterns      =         4
Prior: uniform                           Obs per pattern: min =        44
                                                          avg =    568.75
                                                          max =      1459

Observed log likelihood = -11562.213 at iteration 15

------------------------------------------------------------------------------
             | agessco~d  Treatment  agessco~e  area_cl~a
-------------+----------------------------------------------------------------
Coef
       _cons |   98.0356    .512967   102.8603   2.052308
-------------+----------------------------------------------------------------
Sigma
agesscore_~d |  207.4052  -.1979546   126.8153  -.6649495
   Treatment | -.1979546   .2498319  -.0816383   -.007931
agesscore_~e |  126.8153  -.0816383   196.1873  -.9095955
area_class~a | -.6649495   -.007931  -.9095955   .7352859
------------------------------------------------------------------------------

Little's MCAR test

Number of obs       = 2275
Chi-square distance = 104.5477
Degrees of freedom  = 8
Prob > chi-square   = 0.0000

.
```

Pattern mixture model

```
.        xi: rctmiss, pmmdelta(-1.5): reg agesscore_PIRA_end i.Treatment_archived agesscore_PIRA_base i.area_classes_strata if agesscore_PIRA_base!=., cluste
> r(MOU_Number)
i.Treatment_a~d   _ITreatment_0-1     (naturally coded; _ITreatment_0 omitted)
i.area_classe~a   _Iarea_clas_1-4     (naturally coded; _Iarea_clas_1 omitted)

---------- RCT analysis allowing for informatively missing outcomes ----------
Observed outcomes:       1459 (66 clusters)
Unobserved outcomes:     716 (80 clusters)
Missing data assumption: MNAR
Missing data model:      Pattern-mixture model
Delta:                   -1.5
Auxiliary variables:     (none)
Estimation method:       Two linear regressions
Variances clustered on:  MOU_Number
Effective sample size:   1463.8247 (66.170581 clusters)
------------------------------------------------------------------------------
 agesscore_PIRA_end | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
--------------------+---------------------------------------------------------
      _ITreatment_1 |  -.1929509    1.42099    -0.14   0.892    -3.030723    2.644821
agesscore_PIRA_base |   .6453904   .0285141    22.63   0.000     .5884466    .7023341
      _Iarea_clas_2 |  -.7926691   1.608808    -0.49   0.624    -4.005522    2.420184
      _Iarea_clas_3 |  -.7177412   1.662029    -0.43   0.667    -4.036879    2.601396
      _Iarea_clas_4 |   .2001608   2.028215     0.10   0.922    -3.850264    4.250586
              _cons |   31.76053    3.18987     9.96   0.000     25.39024    38.13082
------------------------------------------------------------------------------

.
```

Subgroup analysis

FSM Interaction term

```
.                xtmixed agesscore_PIRA_end ib2.interaction_FSM_treatment agesscore_PIRA_base i.area_classes_strata || MOU_Number: if agesscore_PIRA_base!=. & agesscore_PIRA_e
> nd!=., mle var

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log likelihood = -5492.3696
Iteration 1:   log likelihood = -5492.3696

Computing standard errors:

Mixed-effects ML regression                     Number of obs     =      1,452
Group variable: MOU_Number                      Number of groups  =         66

                                                Obs per group:
                                                              min =          5
                                                              avg =       22.0
                                                              max =         51

                                                Wald chi2(7)      =    1014.19
Log likelihood = -5492.3696                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
       agesscore_PIRA_end |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
--------------------------+---------------------------------------------------
interaction_FSM_treatment |
          Control(BAU) 0  |   3.650832   1.038917     3.51   0.000     1.614592    5.687072
                 PALS 0   |    2.28247   1.409908     1.62   0.105    -.4808983    5.045838
                 PALS 1   |   .7383321   1.524524     0.48   0.628     -2.24968    3.726344
                          |
      agesscore_PIRA_base |   .6325115   .0208471    30.34   0.000     .5916519     .673371
                          |
      area_classes_strata |
       Midlands >1 class  |  -.7127871   1.363277    -0.52   0.601    -3.384761    1.959187
      North East 1 class  |  -.8200085   1.536329    -0.53   0.594    -3.831158    2.191141
     North East >1 class  |  -.0068528   2.260663    -0.00   0.998    -4.437672    4.423966
                          |
                   _cons  |   31.18659   2.539064    12.28   0.000     26.21012    36.16307
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
MOU_Number: Identity         |
                  var(_cons) |   15.49217   3.524583      9.918767     24.1973
-----------------------------+------------------------------------------------
               var(Residual) |   106.0261   4.021909      98.42924    114.2094
------------------------------------------------------------------------------
LR test vs. linear model: chibar2(01) = 119.98        Prob >= chibar2 = 0.0000
```

FSM only

```
.                xtmixed agesscore_PIRA_end Treatment agesscore_PIRA_base i.area_classes_strata || MOU_Number: if FSM_ANALYSIS==1 & agesscore_PIRA_base!=. & agesscore_PIRA_end
> !=., mle var

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log likelihood =  -1400.374
Iteration 1:   log likelihood = -1400.3739
Iteration 2:   log likelihood = -1400.3739

Computing standard errors:

Mixed-effects ML regression                     Number of obs     =        367
Group variable: MOU_Number                      Number of groups  =         61

                                                Obs per group:
                                                              min =          1
                                                              avg =        6.0
                                                              max =         15

                                                Wald chi2(5)      =     171.27
Log likelihood = -1400.3739                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
  agesscore_PIRA_end |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------------+--------------------------------------------------------
           Treatment |   1.969476   1.663446     1.18   0.236    -1.290818    5.229769
 agesscore_PIRA_base |    .575228   .0446538    12.88   0.000     .4877082    .6627477
                     |
 area_classes_strata |
  Midlands >1 class  |  -.9684807   2.080405    -0.47   0.642       -5.046    3.109039
 North East 1 class  |   .5401261   2.198339     0.25   0.806    -3.768539    4.848791
North East >1 class  |  -1.008382   3.374015    -0.30   0.765     -7.62133    5.604566
                     |
               _cons |   35.65303   4.926119     7.24   0.000     25.99801    45.30804
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
MOU_Number: Identity         |
                  var(_cons) |   17.79926   6.693158      8.517652    37.19497
-----------------------------+------------------------------------------------
               var(Residual) |   108.6449   8.626966       92.9864    126.9402
------------------------------------------------------------------------------
LR test vs. linear model: chibar2(01) = 18.36         Prob >= chibar2 = 0.0000
```

EAL interaction

```
. xtmixed agesscore_PIRA_end ib2.INTERACTION_EAL_TREAT agesscore_PIRA_base i.area_classes_strata || MOU_Number: if agesscore_PIRA_base!=. & agesscore_PIRA_end!=
> ., mle var

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log likelihood = -5494.4256
Iteration 1:   log likelihood = -5494.4256

Computing standard errors:

Mixed-effects ML regression                     Number of obs     =      1,451
Group variable: MOU_Number                      Number of groups  =         66

                                                Obs per group:
                                                              min =          5
                                                              avg =       22.0
                                                              max =         51

                                                Wald chi2(7)      =     994.77
Log likelihood = -5494.4256                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
   agesscore_PIRA_end |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------------------+-------------------------------------------------------
INTERACTION_EAL_TREAT |
      Control(BAU) 0  |  -.6725569   1.431411    -0.47   0.638    -3.478071    2.132957
             PALS 0   |  -1.830753   1.714578    -1.07   0.286    -5.191264    1.529758
             PALS 1   |   .6255425   1.975699     0.32   0.752    -3.246756    4.497841
                      |
  agesscore_PIRA_base |   .6482974    .020669    31.37   0.000     .6077869    .6888078
                      |
  area_classes_strata |
   Midlands >1 class  |  -.9962776   1.358772    -0.73   0.463    -3.659422    1.666867
  North East 1 class  |  -1.045531   1.529111    -0.68   0.494    -4.042533    1.951471
 North East >1 class  |  -.1380531   2.248554    -0.06   0.951    -4.545138    4.269032
                      |
               _cons  |   33.02486   2.767844    11.93   0.000     27.59999    38.44974
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
MOU_Number: Identity         |
                  var(_cons) |   15.23557   3.523721      9.682568    23.97324
-----------------------------+------------------------------------------------
               var(Residual) |   106.9717   4.061033      99.30115    115.2348
------------------------------------------------------------------------------
LR test vs. linear model: chibar2(01) = 110.67        Prob >= chibar2 = 0.0000
```

EAL only

```
. xtmixed agesscore_PIRA_end Treatment agesscore_PIRA_base i.area_classes_strata || MOU_Number: if EAL_ANALYSIS==1 & agesscore_PIRA_base!=. & agesscore_PIRA_end
> !=., mle var

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log likelihood = -880.43126
Iteration 1:   log likelihood = -880.29151
Iteration 2:   log likelihood = -880.29148

Computing standard errors:

Mixed-effects ML regression                     Number of obs     =        228
Group variable: MOU_Number                      Number of groups  =         33

                                                Obs per group:
                                                              min =          1
                                                              avg =        6.9
                                                              max =         26

                                                Wald chi2(5)      =     137.63
Log likelihood = -880.29148                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
  agesscore_PIRA_end |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------------+--------------------------------------------------------
           Treatment |    .188177   1.886277     0.10   0.921    -3.508858    3.885212
 agesscore_PIRA_base |   .6187084   .0539244    11.47   0.000     .5130186    .7243983
                     |
 area_classes_strata |
  Midlands >1 class  |   2.510941   1.875604     1.34   0.181    -1.165176    6.187058
 North East 1 class  |   6.290445   5.132694     1.23   0.220     -3.76945    16.35034
North East >1 class  |   -.611619       4.07    -0.15   0.881    -8.588672    7.365434
                     |
               _cons |   35.34239   5.806226     6.09   0.000     23.96239    46.72238
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
MOU_Number: Identity         |
                  var(_cons) |   2.481712   4.760886      .0577852    106.5826
-----------------------------+------------------------------------------------
               var(Residual) |   129.9203   12.77324      107.1496    157.5302
------------------------------------------------------------------------------
LR test vs. linear model: chibar2(01) = 0.35          Prob >= chibar2 = 0.2768
```

Pupils with lower reading attainment at baseline

```
.         xtmixed agesscore_PIRA_end ib2.interaction_ave_PIRA_base_treat agesscore_PIRA_base i.area_form_strata || sch_harmonised_num: if agesscore_PIRA_base!=. & a
> gesscore_PIRA_end!=.

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log likelihood =  -5527.729
Iteration 1:   log likelihood =  -5527.729

Computing standard errors:

Mixed-effects ML regression                     Number of obs     =      1,459
Group variable: sch_harmonis~m                  Number of groups  =         66
                                                Obs per group:
                                                              min =          5
                                                              avg =       22.1
                                                              max =         51
                                                Wald chi2(7)      =     994.96
Log likelihood =  -5527.729                     Prob > chi2       =     0.0000
```

agesscore_PIRA_end	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
interaction_ave_PIRA_base_treat						
Control(BAU) Average or above	.0224825	1.174142	0.02	0.985	-2.278793	2.323758
PALS Average or above	-.624113	1.52661	-0.41	0.683	-3.616213	2.367987
PALS Below average	-1.1527	1.397538	-0.82	0.409	-3.891825	1.586425
agesscore_PIRA_base	.6381128	.0327724	19.47	0.000	.57388	.7023455
area_form_strata						
Midlands/>1 class	-.8880125	1.400882	-0.63	0.526	-3.633691	1.857666
North East/1 class	-1.165139	1.571801	-0.74	0.459	-4.245813	1.915534
North East/>1 class	-.0582976	2.313629	-0.03	0.980	-4.592926	4.476331
_cons	33.4632	3.255186	10.28	0.000	27.08315	39.84325

Random-effects parameters	Estimate	Std. err.	[95% conf.	interval]
sch_harmon~m: Identity				
sd(_cons)	4.071575	.4577783	3.266326	5.075342
sd(Residual)	10.35032	.1958504	9.973486	10.74138

LR test vs. linear model: chibar2(01) = 128.07 Prob >= chibar2 = 0.0000

Pupils with higher reading attainment at baseline

```
.         xtmixed agesscore_PIRA_end ib1.interaction_ave_PIRA_base_treat agesscore_PIRA_base i.area_form_strata || sch_harmonised_num: if agesscore_PIRA_base!=. & a
> gesscore_PIRA_end!=.

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log likelihood =  -5527.729
Iteration 1:   log likelihood =  -5527.729

Computing standard errors:

Mixed-effects ML regression                     Number of obs     =      1,459
Group variable: sch_harmonis~m                  Number of groups  =         66
                                                Obs per group:
                                                              min =          5
                                                              avg =       22.1
                                                              max =         51
                                                Wald chi2(7)      =     994.96
Log likelihood =  -5527.729                     Prob > chi2       =     0.0000
```

agesscore_PIRA_end	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
interaction_ave_PIRA_base_treat						
Control(BAU) Below average	-.0224825	1.174142	-0.02	0.985	-2.323758	2.278793
PALS Average or above	-.6465955	1.261625	-0.51	0.608	-3.119335	1.826143
PALS Below average	-1.175182	1.510131	-0.78	0.436	-4.134985	1.784621
agesscore_PIRA_base	.6381128	.0327724	19.47	0.000	.57388	.7023455
area_form_strata						
Midlands/>1 class	-.8880125	1.400882	-0.63	0.526	-3.633691	1.857666
North East/1 class	-1.165139	1.571801	-0.74	0.459	-4.245813	1.915534
North East/>1 class	-.0582976	2.313629	-0.03	0.980	-4.592926	4.476331
_cons	33.48568	3.895522	8.60	0.000	25.8506	41.12076

Random-effects parameters	Estimate	Std. err.	[95% conf.	interval]
sch_harmon~m: Identity				
sd(_cons)	4.071575	.4577783	3.266326	5.075342
sd(Residual)	10.35032	.1958504	9.973486	10.74138

LR test vs. linear model: chibar2(01) = 128.07 Prob >= chibar2 = 0.0000

Pupils with very low reading attainment at baseline

```
.         xtmixed agesscore_PIRA_end ib1.interaction_quartiles_base_treat agesscore_PIRA_base i.area_form_strata || sch_harmonised_num: if agesscore_PIRA_base!=. &
> agesscore_PIRA_end!=.

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log likelihood = -5525.1506
Iteration 1:   log likelihood = -5525.1506

Computing standard errors:

Mixed-effects ML regression                     Number of obs     =      1,459
Group variable: sch_harmonis~m                  Number of groups  =         66
                                                Obs per group:
                                                              min =          5
                                                              avg =       22.1
                                                              max =         51
                                                Wald chi2(11)     =    1003.64
Log likelihood = -5525.1506                     Prob > chi2       =     0.0000
```

agesscore_PIRA_end	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
interaction_quartiles_base_treat						
Control(BAU) Q2	-.3485483	1.444296	-0.24	0.809	-3.179316	2.48222
Control(BAU) Q3	-.6135113	1.745195	-0.35	0.725	-4.034031	2.807008
Control(BAU) Q4	1.171364	2.280163	0.51	0.607	-3.297674	5.640402
PALS Q1	-2.025447	1.542019	-1.31	0.189	-5.047749	.9968543
PALS Q2	-.6365585	1.697797	-0.37	0.708	-3.96418	2.691063
PALS Q3	-1.304295	1.980954	-0.66	0.510	-5.186893	2.578303
PALS Q4	.7913134	2.447113	0.32	0.746	-4.00494	5.587566
agesscore_PIRA_base	.6032971	.053864	11.20	0.000	.4977256	.7088686
area_form_strata						
Midlands/>1 class	-.843917	1.400635	-0.60	0.547	-3.58911	1.901276
North East/1 class	-1.194334	1.572027	-0.76	0.447	-4.275451	1.886782
North East/>1 class	-.153516	2.31339	-0.07	0.947	-4.687677	4.380645
_cons	37.03558	4.841621	7.65	0.000	27.54618	46.52498

Random-effects parameters	Estimate	Std. err.	[95% conf.	interval]
sch_harmon~m: Identity				
sd(_cons)	4.070637	.4580108	3.265042	5.074999
sd(Residual)	10.33148	.1955069	9.955311	10.72186

LR test vs. linear model: chibar2(01) = 127.60 Prob >= chibar2 = 0.0000

Pupils with very high reading attainment at baseline

```
.         xtmixed agesscore_PIRA_end ib4.interaction_quartiles_base_treat agesscore_PIRA_base i.area_form_strata || sch_harmonised_num: if agesscore_PIRA_base!=. &
> agesscore_PIRA_end!=.

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log likelihood = -5525.1506
Iteration 1:   log likelihood = -5525.1506

Computing standard errors:

Mixed-effects ML regression                     Number of obs     =      1,459
Group variable: sch_harmonis~m                  Number of groups  =         66
                                                Obs per group:
                                                              min =          5
                                                              avg =       22.1
                                                              max =         51
                                                Wald chi2(11)     =    1003.64
Log likelihood = -5525.1506                     Prob > chi2       =     0.0000
```

agesscore_PIRA_end	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
interaction_quartiles_base_treat						
Control(BAU) Q1	-1.171364	2.280163	-0.51	0.607	-5.640402	3.297674
Control(BAU) Q2	-1.519912	1.694177	-0.90	0.370	-4.840437	1.800613
Control(BAU) Q3	-1.784875	1.367562	-1.31	0.192	-4.465248	.8954977
PALS Q1	-3.196811	2.483359	-1.29	0.198	-8.064105	1.670483
PALS Q2	-1.807922	1.903421	-0.95	0.342	-5.538559	1.922715
PALS Q3	-2.475659	1.678252	-1.48	0.140	-5.764973	.8136547
PALS Q4	-.3800505	1.552791	-0.24	0.807	-3.423465	2.663364
agesscore_PIRA_base	.6032971	.053864	11.20	0.000	.4977256	.7088686
area_form_strata						
Midlands/>1 class	-.843917	1.400635	-0.60	0.547	-3.58911	1.901276
North East/1 class	-1.194334	1.572027	-0.76	0.447	-4.275451	1.886782
North East/>1 class	-.153516	2.31339	-0.07	0.947	-4.687677	4.380645
_cons	38.20694	6.673867	5.72	0.000	25.12641	51.28748

Random-effects parameters	Estimate	Std. err.	[95% conf.	interval]
sch_harmon~m: Identity				
sd(_cons)	4.070637	.4580108	3.265042	5.074999
sd(Residual)	10.33148	.1955069	9.955311	10.72186

LR test vs. linear model: chibar2(01) = 127.60 Prob >= chibar2 = 0.0000

Subgroup analysis: FSM and EAL sample effect size estimations

FSM pupils only

Table 48 Sub-group analysis: FSM pupils only

			Intervention group: FSM Subgroup		Control group: FSM Subgroup		
Outcome	Unadjusted differences in means	Adjusted differences in means (95% CI)	n (missing)	Variance of outcome	n (missing)	Variance of outcome	Pooled variance
Primary outcome: FSM group	2.269	1.969 (-1.291 - 5.230)	220 (62)	187.47	147 (144)	188.50	188.6108

Table 49 Effect size calculation: sub-group analysis: FSM pupils only

	Unadjusted means				Effect size		
	Intervention group: FSM subgroup		Control group: FSM subgroup				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention; control)	Hedges g (95% CI)	p-value
Primary outcome: FSM group	220 (68)	93.991 (92.171; 95.810)	147 (150)	91.721 (89.483; 93.959)	367 (220, 147)	0.144 (-0.09; 0.38)	0.236

EAL pupils only

Table 50 Sub-group analysis: EAL pupils

			Intervention group: EAL Subgroup		Control group: EAL Subgroup		
Outcome	Unadjusted differences in means	Adjusted differences in means (95% CI)	n (missing)	Variance of outcome	n (missing)	Variance of outcome	Pooled variance

Primary outcome: EAL group	2.075	0.188 (-3.509 - 3.885)	156 (30)	219.32	72 (103)	210.39	216.4964

Table 51 Effect size calculation: sub-group analysis: EAL group

	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
Primary outcome: EAL group	156 (31)	98.244 (95.901; 100.586)	72 (105)	100.319 (96.911; 103.728)	228 (156,72)	0.013(-0.24; 0.26)	0.92

Appendix H: Implementation and Process Evaluation Privacy Notices

Privacy notice for the surveys

Privacy notice

RAND Europe is collecting data on the implementation of the Peer Assisted Learning Strategies for Reading UK (PALS-UK) on the basis of legitimate interest as we have been contracted by the Education Endowment Foundation to evaluate this intervention and you have agreed to take part in this survey. Because the survey will ask you to provide the name of your school and your role in the school, your responses to the survey will amount to identifiable data.

Your survey responses will be collected and stored on the SmartSurvey platform by RAND Europe. RAND Europe will obtain the data securely from SmartSurvey. SmartSurvey will delete your survey responses and identifiable data once RAND Europe has obtained it. RAND Europe will maintain this data in confidence and use it only for the purpose of evaluating PALS-UK. The data will be stored securely on RAND Europe's data servers for the duration of the PALS-UK evaluation project. To allow us time to analyse and report the results of the trial, this period will extend beyond your school's participation in the programme. Your responses will be used to create descriptive statistics and individual schools will not be identified in this context. Your responses shall not be made available to your employer or otherwise passed to any third party.

Please do not provide any sensitive data in this survey, such as your political persuasion. If sensitive data is provided in the survey, RAND Europe will delete it before analysis.

In certain circumstances, you may have the right to restrict or object to processing. You also have the right to make a subject access request to see all the information held about you. To exercise any of these rights, please contact the RAND Europe data protection officer (**redpo@randeurope.org**). If you have any questions about how your data will be used, please do not hesitate to contact the RAND Europe data protection officer (**redpo@randeurope.org**). You may also contact the UK Information Commissioner's Office if you have any concerns about our use of your data at https://ico.org.uk/concerns/.

Please click 'Next Page' if you would like to proceed.

Privacy notice for interviews

Privacy Notice
PALS UK evaluation case studies

About the project, who we are and what data we collect

The accompanying information sheets outline information about the project. This privacy notice outlines how your data will be used as part of the project.

RAND Europe Community Interest Company is a not-for-profit research organisation registered in the UK conducting independent research to inform policy.

In this project we will collect your name, email and telephone number.

Why are we collecting it?

We are collecting your data in order to arrange your participation in a telephone interview. These interviews make up part of the research activity being undertaken to evaluate the impact of the Peer Assisted Learning Strategies UK (PALS-UK).

What is the legal basis for processing your data?

We are using your data on the basis of our legitimate interests. Your data is collected and processed solely to facilitate your voluntary contribution to the project. The data is not excessive and will be used for the purposes of contacting you to arrange your participation in the interview, as required to meet the project goals. These project goals have been explained to you and will lead to a wider public benefit through our work. The data is necessary for the purpose of the project, as without we would be unable to undertake the interview or recognise your contribution. We also judge that there is very limited scope for harm to you as appropriate data handling safeguards have been put in place. As such the approach to processing balances our legitimate interests against your interests, rights and freedoms.

What do we use the data for?

We will use your data to contact you. We will use a random unique ID to attribute any contribution of yours that is used in our report, not your name.

How do we share the data, and how do we keep your data secure?

We will keep all data safe on our secure servers. We will not share your data with any third parties.

How long do we keep your data?

Your data will be deleted within 12 months of the end of the project (end of project estimated December 2020).

What choices do you have in our use of your data?

You may contact us to request the deletion of your personal data.

What are your rights?

RAND Europe operates in accordance with the Data Protection Act 2018 and EU law including GDPR. You are provided with certain rights that you may have the right to exercise through us. In summary those rights are:

- To access, correct or erase your data. Your right to erase your name in relation to any attribution shall expire after it has been submitted for publication.
- To object to the processing of your data. Your right to object to processing of your name in relation to any attribution shall expire after it has been submitted for publication.
- To request that our processing or your data is restricted. Your right to restrict processing of your name in relation to any attribution shall expire after it has been submitted for publication.

If you wish to exercise any of these rights please contact the RAND Europe Data Protection Officer by email at **REdpo@rand.org** or in writing to Data Protection Officer, RAND Europe, Westbrook Centre, Milton Road, Cambridge, CB4 1YG, UK.

How do you contact us?

You can contact us by email at **PALS@randeurope.org**.

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

You may re-use this document/publication (not including logos) free of charge in any format or medium, under the terms of the Open Government Licence v3.0.

To view this licence, visit **https://nationalarchives.gov.uk/doc/open-government-licence/version/3** or email: **psi@nationalarchives.gsi.gov.uk**

Where we have identified any third-party copyright information you will need to obtain permission from the copyright holders concerned. The views expressed in this report are the authors' and do not necessarily reflect those of the Department for Education.

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

Appendix I: Secondary outcome - Reading comprehension (WIAT-III-UK-T)

As discussed in the updated evaluation protocol and SAP, the WIAT-III-UK-T element of the outcome testing was only partially administered. Indeed, data for only 18 schools is available for the WIAT-III-UK reading comprehension subtest; 11 of these were control schools, while 7 were PALS-UK schools. None of the schools completed the WIAT-III-UK-T reading fluency subtest. Owing to the COVID 19 disruptions (see Changes to the evaluation objectives), a secondary outcome analysis on pupil's oral reading fluency and reading comprehension using the WIAT-III-UK-T subtests is not possible. We report mean scores on the Reading comprehension WIAT-III-UK-T element for the small number of schools (n=18).

It was intended in the design of the secondary outcome analysis to mirror the primary outcome analysis where relevant. To this end, we opted to take the age standardised scores on the WIAT-III-UK-T reading comprehension subtest as the secondary outcome measure of interest for this subtest. In total, endline data for the WIAT-III-UK-T reading comprehension subtest was available for 163 pupils.

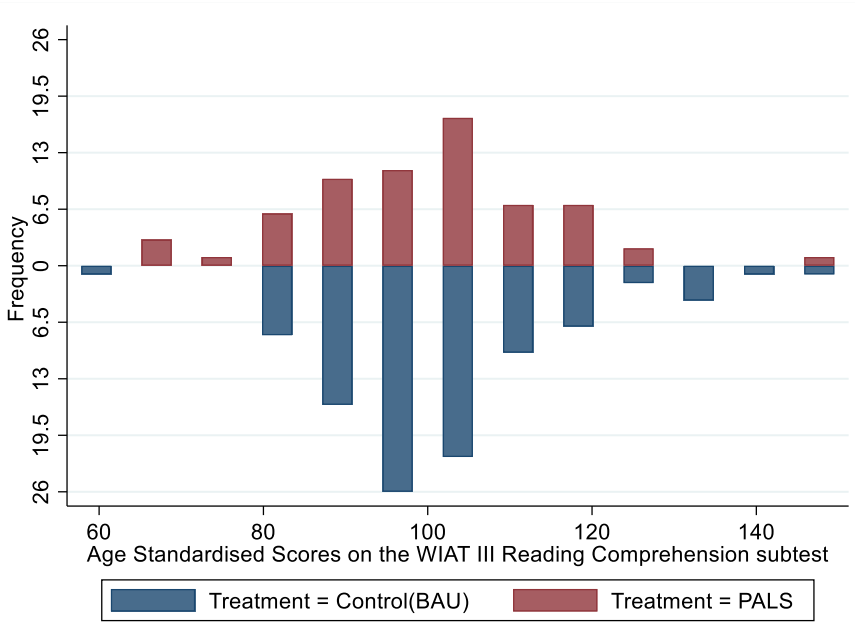
For this overall sample, the mean age standardised score on the subtest was 100.264 and the standard deviation of 14.678. Table 52 presents summary statistics on this secondary outcome measure by treatment condition. The number of pupils in the sample was higher in the control group (98) as compared to the intervention group (65).

Table 52 WIAT-III-UK-T Reading Comprehension subtest mean scores and sample sizes by treatment allocation

	Intervention group		Control group		
Secondary measure	n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)	Difference in raw means (I-C)
WIAT-III-UK-T Reading Comprehension subtest	65	99.415 (14.701)	98	100.827 (14.712)	-1.412

The mean age standardised score on this subtest was slightly lower among pupils in the intervention group (99.415) as compared to pupils in the control group (100.827), although the difference was minimal (-1.412). The distribution of the age standardised scores on this subtest by treatment condition is shown in Figure 33.

Figure 36 Distribution of age standardised scores on the WIAT-III-UK-T Reading Comprehension subtest at endline by treatment condition



Appendix J: Full observation data

Case Study ID	Visit	% correct: CSU	% correct: PRR	% correct: PS	% correct: PR	% correct: CM	Unweighted overall % correct	Weighted overall % correct
School 1	Peer Observation 1	100%	86%	86%	N/A	100%	93%	89%
	Peer Observation 2	100%	86%	86%	91%	100%	93%	89%
	Peer Observation 3	100%	86%	82%	95%	100%	93%	89%
	RA observation 1	100%	91%	N/A	N/A	100%	97%	95%
	RA observation 2	100%	86%	77%	91%	100%	91%	86%
School 2	Peer Observation 1	100%	100%	100%	100%	100%	100%	100%
	Peer Observation 2	100%	93%	95%	100%	100%	98%	97%
	Peer Observation 3	100%	100%	100%	100%	100%	100%	100%
	RA observation 1	100%	100%	N/A	N/A	100%	100%	100%
	RA observation 2	100%	100%	95%	95%	100%	98%	97%
School 3	RA observation 1	100%	93%	N/A	N/A	100%	98%	95%
	RA observation 2	100%	71%	91%	100%	100%	92%	91%
School 4	Peer Observation 1	100%	N/A	N/A	95%	100%	98%	97%
	Peer Observation 2	100%	64%	100%	95%	100%	92%	91%
	Peer Observation 3	100%	71%	91%	100%	100%	92%	91%
	RA observation 1	100%	100%	N/A	N/A	100%	100%	100%
	RA observation 2	100%	100%	86%	100%	100%	97%	95%
School 5	RA observation 1	100%	93%	82%	100%	100%	95%	92%
	RA observation 2	100%	100%	77%	N/A	100%	94%	89%
School 6	Peer Observation 1	100%	100%	100%	91%	100%	98%	97%
	RA observation 1	100%	93%	100%	N/A	N/A	98%	45%
	RA observation 2	100%	71%	95%	73%	100%	88%	83%
School 7	Peer Observation 1	100%	100%	100%	100%	100%	100%	100%
	Peer Observation 2	100%	86%	100%	100%	100%	97%	97%
	Peer Observation 3	100%	100%	100%	100%	100%	100%	100%
	Peer Observation 4	100%	100%	100%	100%	100%	100%	100%
	RA observation 1	100%	71%	N/A	N/A	100%	90%	82%
	RA observation 2	N/A	93%	91%	95%	100%	95%	93%

School 8	Peer Observation 1	100%	93%	32%	70%	100%	79%	66%
	Peer Observation 2	100%	100%	100%	100%	100%	100%	100%
	RA observation 1	100%	100%	N/A	N/A	100%	100%	100%
	RA observation 2	100%	100%	95%	89%	100%	97%	95%

Note: N/A represents cases where all activities within themes were not observed as training was instead observed, so a fidelity score could not be calculated. Abbreviations are as follows: CSU=Classroom set-up; PRR=Partner reading with re-tell; PS=Paragraph shrinking; PR=Prediction relay; CM=Classroom management. Number of components per theme is as follows: CSU, n=6; PRR, n=14; PS, n=22; PR, n=22; CM, n=3.