



Lexia Reading Core5®

Evaluation Report

September 2021

Updated February 2022

Louise Tracey, Louise Elliott, Caroline Fairhurst, Laura Mandefield,
Imogen Fountain, Sarah Ellison



UNIVERSITY
of York







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

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

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

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

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

-  Jonathan Kay
Education Endowment Foundation
5th Floor, Millbank Tower
21–24 Millbank
SW1P 4QP
-  0207 802 1653
-  jonathan.kay@eefoundation.org.uk
-  www.educationendowmentfoundation.org.uk



Contents

About the evaluator	3
Executive summary	4
Introduction	6
Methods.....	13
Impact evaluation.....	25
Conclusion.....	63
References	67
Appendix A: EEF cost rating.....	69
Appendix B: Security classification of trial findings.....	70
Appendix C: Changes since the previous evaluation	72
Appendix D: Effect size estimation	73

About the evaluator

The project was independently evaluated by a team from the University of York: Louise Tracey, Louise Elliott, Caroline Fairhurst, Laura Mandefield, Imogen Fountain, and Sarah Ellison.

The principal investigators were Dr Louise Tracey and Louise Elliott.

Contact details:

Dr Louise Tracey
Senior Research Fellow
Department of Education
University of York
YO10 5DD

Email: louise.tracey@york.ac.uk

Tel: 01904 328160

Acknowledgements

With many thanks to all the schools, teachers, and children who participated in the study. Your contribution has been invaluable. Also thank you to all the data collectors and data inputters and to Saleema Rex, who contributed to the writing of the statistical analysis plan. Especial mention to the team at LexiaUK (led by Rob Kay and Norman Robinson) and the team at Queen's University, Belfast (Dr Maria Cockerill, Dr Joanne O'Keeffe, and Professor Allen Thurston).

Executive summary

The project

Lexia Reading Core5¹ (Lexia) is a computer-based integrated learning system that aims to improve reading skills developed by Lexia Learning Systems LLC. It consists of three elements: personalised online student activities, real-time reporting of student progress, and paper-based resources to guide instruction where needed. Teachers can use it to target struggling readers as a whole-class or whole-school intervention or as a home-use supplement to teaching.

In this study, pupils began with an online diagnostic assessment that placed them at the correct starting point. They then worked independently in small groups away from their classroom setting. Facilitators, who are teachers or teaching assistants (TAs), supported pupils in these sessions. The facilitators gave guidance on how to use the software, reinforced some units, and tracked pupil progress. This project aimed to evaluate the efficacy of Lexia in improving the reading ability of struggling readers in Year 2 (children aged six to seven). Independent assessors tested children and those with the lowest scores were selected for the trial. Schools were asked to schedule four sessions of 30 minutes per week for those children (including ten minutes set-up time) for 12 to 24 weeks. Groups of six or seven children took part in each session. Queen's University Belfast provided the delivery team. LexiaUK provided support to schools in the trial.

Lexia was evaluated using a two-armed, randomised controlled efficacy trial. In each school, half of the struggling readers were randomly selected to receive Lexia and the others received their regular lessons and support. Fifty-seven schools and 697 pupils participated in the trial. The schools were from the North East, Wiltshire, Lancashire, London, and Yorkshire and Humber. The trial ran from September 2018 to July 2019. Attainment was measured by a reading test. The implementation and process evaluation included teacher surveys, interviews, and pupil focus groups.

The previous version of this report (published in September 2021) was published without the FSM subgroup analysis due to delays in accessing data from the National Pupil Database (NPD). This updated version incorporates this analysis.

Table 1: Key conclusions

Key conclusions
1. Children offered Lexia made, on average, the equivalent of two additional months' progress in reading compared to other children. This result has a high security rating.
2. Exploratory analysis suggests that children offered Lexia made the equivalent of two additional months' progress in word recognition and decoding skills and one additional month of progress in reading fluency and comprehension skills, on average, compared to other children. Children offered Lexia made, on average, no additional progress in Key Stage 1 national test reading raw scores compared to other children.
3. Children eligible for Free School Meals (FSM) who were offered Lexia made, on average, the equivalent of three additional months' progress in reading compared to other children eligible for FSM. However, this result has high statistical uncertainty.
4. Implementation fidelity was high. Most schools incorporated Lexia into their routine so that pupils received the normal provision for struggling readers as well as this intervention. Over three-quarters of survey respondents were satisfied with the access to the online activities, support provided, and the usefulness of the reports.
5. Most pupils managed to work independently with little scaffolding from teaching assistants. However, teachers and teaching assistants felt that Lexia was unsuitable for a small number of pupils, especially those with Special Educational Needs and Disabilities and those whose reading level was very low.

EEF security rating

These findings have a high security rating. This was an efficacy trial, which tested whether the intervention worked under developer-led conditions in a number of schools. The trial was a well-designed and well-powered two-armed randomised controlled trial. Eleven percent of the pupils who started the trial were not included in the final analysis because some pupils had left the schools in the trial and some pupils had incomplete assessment data. The pupils randomly selected to take part in Lexia sessions were similar to those not selected in terms of prior attainment.

¹ 'Lexia Reading Core5' is a registered trademark of Lexia Learning Systems LLC.

Additional findings

Children offered Lexia made, on average, two additional months' progress compared to those in the control group equivalent. This is our best estimate of impact, which has a high security rating. As with any study, there is always some uncertainty around the result: the possible impact of this programme also includes zero months of additional progress and positive effects of up to four months of additional progress. In terms of secondary outcomes, children offered Lexia made, on average, two additional months' progress in the Word Identification and Word Attack subscales compared to the control group equivalent. These are key building blocks for this age for literacy development. Children offered Lexia also made one month's additional progress in comprehension and fluency skills compared to the control group equivalent. Those skills are important for further literacy development. However, there was no observed additional progress in the secondary outcome for KS1 raw reading scores.

Children eligible for Free School Meals (FSM) who were offered Lexia made, on average, the equivalent of three additional months' progress in reading compared to other children eligible for FSM (with the range of possible effects ranging from no additional progress to six months' additional progress).

Of the survey respondents who delivered Lexia, 92% said that the training gave them sufficient information to enable them to deliver it. The same percentage reported being satisfied or very satisfied with the webinars. In most schools, TAs facilitated Lexia sessions. The average number of pupils in a session was six, as expected at the outset. Under normal conditions Lexia can be delivered to more pupils, and for struggling readers of this age up to 15 is typical. This means that children in this trial may have had more adult input than they would have had in a larger group.

Teachers and TAs delivering the intervention used all three components of Lexia though to varying levels. The evaluation defined that a school was compliant if pupils completed at least 60 minutes a week of Lexia sessions for 12 non-consecutive weeks. Approximately 80% of pupils met this target and 93% of pupils used Lexia for at least 720 minutes in total. Regarding the real-time reports, respondents in 92% of schools said they accessed the reports and 69% of respondents said they accessed the reports at least weekly. Not all schools used the paper-based resources regularly. Respondents in interview schools suggest this was due to time constraints and other practicalities.


Some survey respondents said that Lexia was unsuitable for pupils with severe Special Educational Needs or Disabilities or those whose reading levels were very low. The support that those children needed detracted from the interactive and adaptive element of the programme and reduced the time that facilitators had with other pupils. In addition, in a small number of schools TA time was diverted away from the pupils in the control group to those receiving Lexia. Overall, the evaluation supports the existing evidence that Lexia could be effective for struggling readers.

Cost

The estimated cost of Lexia is £3,690 over a three-year period, or £41.00 per pupil per year when averaged over three years. This includes 30 named student licences per year, three one-hour training sessions for school staff in the first year, and ongoing support and training throughout the three-year period. It assumes delivery to six or seven pupils at a time, minimal printing costs, and that schools have available seven devices and headphones to support delivery.

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 skills (WRMT-III composite score)	0.11 (-0.04, 0.27)	2		620 (310; 310)	0.15	£ £ £ £ £
Reading skills, FSM-eligible pupils (WRMT-III composite score)	0.26 (0.01, 0.51)	3	N/A	233 (114; 119)	0.04	N/A

Introduction

Background

In 2019, the percentage of pupils not reaching the expected standard in reading at Key Stage 1 (KS1) was 25%. For pupils in receipt of free school meals (FSM) the figure was 38%. This gap in attainment has remained 'relatively stable' since 2016, when changes were implemented in the national curriculum (DfE, 2019b). Early literacy problems can hinder children's knowledge and development with long-term consequences for their educational outcomes. Evidence from the U.S. suggests that children entering Grade 3 (KS2 in England) still learning to read are likely to continue to struggle academically throughout their school career (Fiester, 2013). Consequently, there is a need for identifying the most promising approaches to support literacy in the early years and KS1 (Higgins, Katsipataki and Coleman, 2014). While it is advocated that approaches to teaching reading in KS1 should integrate phonics, fluency, and comprehension skills at the classroom level, it is also recognised that for some pupils a more targeted approach may be needed using high quality and structured intervention programmes (EEF, 2017). Remedial and tutorial use of technology has been identified as being particularly practical for lower-attaining pupils, those with special educational needs, or those from disadvantaged backgrounds in providing intensive support to enable them to catch up with their peers (Higgins, Xiao and Katsipataki, 2012). Technology can be particularly useful in personalising learning to match pupils' individual abilities and needs given the potential for such programmes to adapt and focus on the child's learning needs. Consequently, programmes implementing an integrated learning system (ILS) approach are popular with schools for instructional purposes. In addition, a review of the effectiveness of educational technology applications in improving the reading achievement of struggling readers in elementary schools in the U.S. suggested that such approaches show promise (Cheung and Slavin, 2012): a blended approach that combines online, computer-based activities with direct teacher instruction has been shown as being particularly effective in improving reading skills (Chambers et al., 2008).

Lexia Reading Core5 is a computer-based independent learning system (ILS) using a blended approach that includes the provision of paper-based resources to supplement learning. It is currently used in over 3,000 schools in the U.K. Previous randomised controlled trials (RCTs) of earlier versions of Lexia in the U.S. found 'potentially positive effects' on alphabets and comprehension but no discernible effects on fluency and general reading achievement (What Works Clearinghouse, 2009). These studies typically took place over a six-month period. A more recent, small-scale longitudinal study in the U.S. (68 participants over three years) indicated that low-performing pupils from kindergarten to Grade 2 showed significant gains after using the programme, although there was no control group (Macaruso et al., 2019) and a further quasi-experimental study suggested the programme was effective in improving the literacy skills of struggling first- and third-graders (Taylor, 2018). As such, this programme shows promise. Although some studies of Lexia have been conducted in the U.K. these have generally been small-scale (Brooks, 2016). One RCT of Lexia in Northern Ireland found positive effects for blending and non-word reading in four- to six-year-old pupils (O'Callaghan et al., 2016). However, there was a small sample size (98 pupils across Reception and Year 1 (equivalence in England) and with only 0.06 to 0.07 effect sizes (0.80 power).

This two-armed efficacy trial, which randomised pupils within schools to receive the Lexia intervention or not, provided an opportunity to evaluate the programme using a large-scale RCT within the U.K. context using outcomes that measure all-round reading ability. The impact evaluation also measured the individual skills that compose 'reading skills', namely word recognition, decoding, comprehension, and fluency, which allows for comparison with previous studies and to identify any areas of strength or weakness in the programme in terms of enhancing children's reading development.

Alongside the impact evaluation, which measures pupil outcomes, an implementation and process evaluation (IPE) was conducted. The IPE was based on the evidence that the effective use of technology in schools is heavily influenced by school and teacher factors (Higgins et al., 2012). This is particularly important given that Lexia is a blended programme utilising personalised online instruction and real-time student data and additional paper-based resources to support this instruction. One small scale study of the programme reinforced this by finding that both the computer time and direct teaching time, when occurring in conjunction, were beneficial for pupils (Ness et al., 2013). Our IPE seeks to provide a useful understanding of the programme and its use in schools including contextual factors for implementation, usage, teacher or teaching assistant scaffolding and support, and any barriers to implementation. In doing so it seeks to inform and explain the findings of the impact evaluation.

Intervention

Lexia, a computer-based ILS, was originally developed in the U.S. to help pupils with dyslexia with their reading. Lexia Reading Core5 is designed for wider use and provides personalised, adaptive learning for a wide range of ability levels at primary school age. It has also been adapted to the U.K. context (e.g. using UK-English audio and spellings). Pupils begin with an online independent diagnostic test and are placed at an appropriate level in the online activities. Pupils then work independently, typically having two to four 20-minute sessions per week (not including set-up time which involves distributing laptops, settling pupils in and log-on to the online element of the program). Facilitators (teachers or teaching assistants) need to give pupils initial guidance on using the online activities, teach and reinforce some units (pupils complete and progress through different units; if they are not deemed to have successfully completed a unit the software adapts, simplifying the activity to provide additional support and instruction and the child repeats the activity to embed their learning)_ and oversee and monitor pupil progression (facilitators are also able to move pupils on if they deem it useful). The system is able to keep track of users' progress in real-time and facilitators are provided with reports to monitor pupils' performance. Where appropriate, paper-based activities, which include scripted lesson plans and independent practise materials for pupils, are suggested from within the system, printed out by the teacher or facilitator. Online training and support is offered to schools using the programme. It is most commonly used in U.K. schools as an in-school supplement to target struggling readers although it can also be used as a whole class intervention or as a school-provided, home-use supplement to teaching.

For this evaluation, schools were issued a package of 30 transferable two-year licences for a two-year period and training by LexiaUK. Each licence was allocated to a particular student. Licences could not be allocated to different students concurrently but could be switched between pupils at the school's discretion for those pupils outside the evaluation year group (Year 1 in 2017/2018, due to be in Year 2 in 2018/2019). For example, if a Year 3 pupil was deemed by the school to have made satisfactory progress the school could then transfer that licence to another pupil in any year group excluding those pupils in Year 2 in 2018/2019. These changes could be made more than once throughout the academic year to different pupils (outside of the evaluation year group and therefore their use was not evaluated). This ability to switch licences was designed in line with usual Lexia practice and also to fit into the wider logic model (Figure 1) where whole-school practice would change. Schools received training: initial training (summer term 2018), reports training (approximately three weeks later) and a progress review (September 2018). Training was delivered via online conference facilities and programme deliverers—teachers and teaching assistants (TAs)—were expected to attend all sessions. In addition, a senior member of staff was also expected to attend the first two training sessions to ensure school buy-in to the programme and wider school support for programme implementation. This was a stepped approach to allow school staff to gain greater knowledge of the programme with a gradual approach. Schools were then expected to practice delivering the programme with their pupils during the summer term (2018) offering the programme to any year group except Year 1 pupils as these could potentially form part of the intervention group during the period of the evaluation. This was to enable facilitators to become familiar with programme delivery and for schools to embed the programme into their existing practice.

From September 2018, schools were expected to deliver the programme to the Year 2 pupils randomised into the intervention condition throughout the academic year and were permitted to use the additional licences with pupils in other year groups (excluding all other Year 2 pupils to avoid confounding factors). For the Year 2 intervention group, however, schools were expected to schedule four sessions of 30 minutes (including ten minutes setup time) per week for, ideally, 24 weeks. To meet the minimum requirements for compliance, pupils were required to do 60 minutes per week (excluding setup time) for at least 12 (not necessarily consecutive) weeks. During the Lexia sessions each pupil was expected to work independently with one adult (either a teacher or TA) supervising the pupils. The adult's role was to ensure the children were on task, monitor progress, scaffold learning with paper-based resources where necessary, and manage the software. It was decided that, for this trial, the intervention should be only delivered in-school as, although some schools currently provide the online element for use at home, it was believed to be important to ensure that pupils without wider access to IT were not disadvantaged. In-school-only provision was also felt to facilitate greater consistency of implementation fidelity. Contamination was unlikely to take place because licences were issued to individual pupils and were monitored by LexiaUK. Please see the TIDieR² table below (Table 3) for further details of the programme as implemented for this efficacy trial.

² Template for Intervention Description and Replication (TIDieR), Hoffmann et al., 2014.

Table 3: TIDieR

Aspect of TIDieR	Exemplification relating to the evaluation
Brief name	Lexia Reading Core5, a computer-based Independent Learning System (ILS) for reading provided by LexiaUK®.
Why: Rationale, theory and/or goal of essential elements of the intervention	The research evidence for the programme is mixed although it does show evidence of promise. In addition, Lexia Reading Core5, and other ILS programmes, are popular with schools for instructional purposes. This evaluation is the first large-scale randomised controlled trial of Lexia Reading Core5 in the U.K.
Who: Recipients of the intervention	Struggling readers in Year 2 identified by the WRMT-R/NU assessment (12 to 14 lowest scoring pupils per class).
What: Physical or informational materials used in the intervention	<p>The following were provided for each school:</p> <ul style="list-style-type: none"> • initial training via online conference facilities; • ongoing support for schools from LexiaUK via online conference facilities or by phone and/or email; • Lexia software; • off-line, paper-based resources including scripted lesson plans and pupil independent-practice materials; and • technical support.
What: Procedures, activities and/or processes used in the intervention	<p>Teachers/TAs trained in the Lexia Reading Core5.³</p> <p>Senior member of staff (such as the headteacher) present at the first two training sessions to encourage school support and implementation.</p> <p>Lexia Reading Core5 accessed by pupils via PC or tablets.</p> <p>Teachers utilise, on a weekly basis, the reports generated by the Lexia software to monitor pupil performance.</p> <p>Where additional need is identified pupils receive paper-based resources and scaffolded teaching either through the facilitator, an additional TA, or the teacher.</p>
Who: Intervention providers/implementers	As an ILS, Lexia Reading Core5 is designed to provide personalised learning to each of the pupils selected. Teachers/TAs facilitated and monitored implementation and provided instruction as required. Teachers/TAs were trained to set up pupils on the Lexia Reading Core5 system and understand the data provided by the online element of the programme. LexiaUK® provided this training virtually via online conference facilities.
How: Mode of delivery	Delivery of Lexia Reading Core5 to the struggling readers in the intervention group, for example, during guided reading sessions, occurred on a PC or tablet computer under the supervision of trained personnel.
Where: Location of the intervention	Schools were advised that children should be withdrawn from the classroom for Lexia Reading Core5 sessions. This was to facilitate the level of teacher/TA monitoring and support required by the programme.
When and how much: Duration and dosage of the intervention	The Lexia Reading Core5 intervention was scheduled for use four times a week for 30 minutes (including ten minutes setup time) over the academic year. For compliance, pupils must have completed a 60 minutes (excluding set-up time) per week for at least 12 weeks (the 12 weeks did not necessarily need to be consecutive).

³ Initial training (summer term 2018), reports training (approximately three weeks later), and a progress review (September 2018).

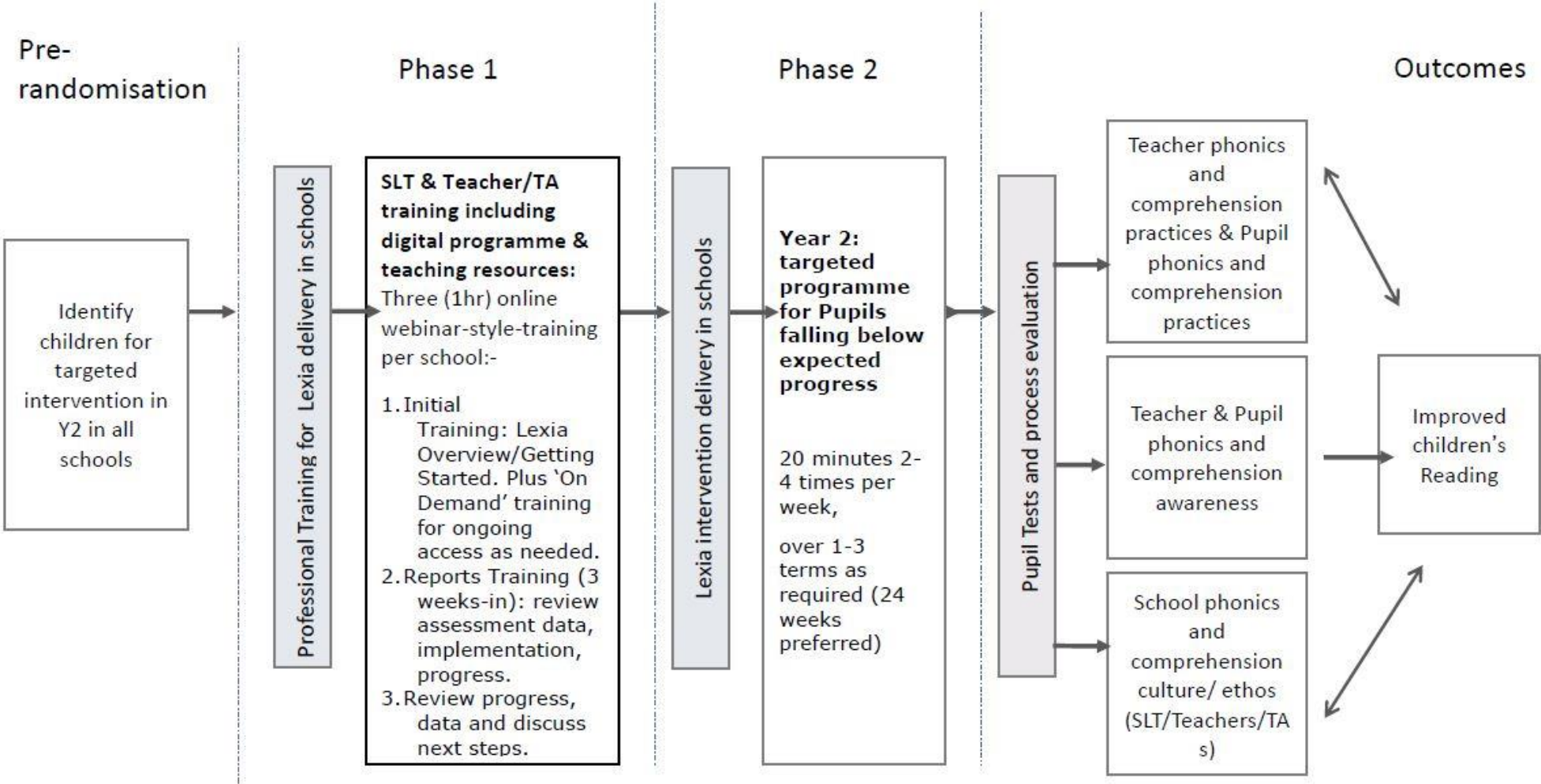
Tailoring: Adaptation of the intervention	Given that the online element of Lexia is an adaptive technology, adaptations were not advised.
How well (planned): Strategies to maximise effective implementation	<p>In order to maximise the effectiveness of the implementation the following strategies were adopted:</p> <ul style="list-style-type: none"> • Teachers/TAs took part in online training sessions; • a member of the school senior leadership should have attended at least the first two of the online training sessions delivered via conference facilities; • ongoing support was provided to facilitating teachers, TAs, and senior leaders; • pupil data was monitored to keep track of compliance and identify those who did not receive the minimum dosage (i.e. non-compliers). LexiaUK® contacted schools via email in the first instance if non-compliance was detected; and • where possible, data from the software was used to assess implementation.

Of the 59 schools that returned a Memorandum of Understanding (MOU), pupils from 57 schools were randomised. Two schools withdrew before randomisation, one due to staffing issues and the other unwilling to sign the Data Sharing Agreement (DSA).

A logic model relating to the evaluation was developed in conjunction with the delivery team and is presented in Figure 1. There were two phases of implementation of the programme. First, training and delivery to children outside of the designated year group to embed delivery (summer 2018). Second, delivery of the programme to the Year 2 intervention pupils (academic year 2018/2019). There were four discrete sections to the evaluation embedded within the logic model:

1. pre-randomisation - assessment and selection of pupils for participation in the evaluation;
2. phase 1 - professional training for Lexia Delivery in schools (please note the logic model focuses on the Year 2 pupils, hence the embedding phase was not included in the logic model);
3. phase 2 - Lexia intervention delivery in schools for Year 2 pupils; as indicated in Figure 1, delivery was intended for 20 minutes a session with a ten-minute set up time (that is, schools were requested to prepare for 30 minutes timetabled delivery allocation per session); and
4. outcomes - measured through the pupil assessments and implementation and process evaluation.

Figure 1: Logic model for evaluation of Lexia Reading Core5®



As described in the TIDieR table (Table 3), inputs included children working individually on the Lexia activities on a computer, with an adult facilitating the sessions, supplemented by paper resources where needed. In terms of outcomes it was anticipated that use of the programme as a whole (including the online and paper-based resources) would impact on both pupil skills and awareness of phonics and comprehension as well as impact on teachers' and teaching assistants' knowledge and practice. It was also felt that, through including wider school staff in the training—such as Year 2 teachers who would oversee the process, senior school managers who would support the use of the programme across the school, as well as TAs who would deliver the programme—and the use of licences for other year groups as the programme became embedded within the school routine, it would also impact across the school in terms of its culture and ethos. Such a move would also result in a scale-up of the programme within schools across year groups providing continuity for pupils. The ultimate outcome, however—as measured by this evaluation—was hypothesised to be improved children's reading skills in the target population (struggling readers in Year 2). In the light of this evaluation, the logic model has been amended and can be found in the Discussion section (Figure 4, p. 63).

Evaluation objectives

The main aim of this evaluation was to assess the impact of Lexia Reading Core5 on struggling readers in Year 2. Consequently, the primary research question was:

1. How effective is Lexia Reading Core5 in improving struggling readers' reading skills during Year 2?

The secondary research questions focus on the effectiveness of the programme on specific reading skills as assessed by national tests (KS1 SATs) and the possible impact on pupils eligible for free school meals (FSM):

2. How effective is Lexia Reading Core5 in improving struggling readers' word recognition skills during Year 2?
3. How effective is Lexia Reading Core5 in improving struggling readers' decoding skills during Year 2?
4. How effective is Lexia Reading Core5 in improving struggling readers' comprehension skills during Year 2?
5. How effective is Lexia Reading Core5 in improving struggling readers' fluency skills during Year 2?
6. How effective is Lexia Reading Core5 in improving struggling readers' outcomes in KS1 national reading assessments?
7. How effective is Lexia Reading Core5 in improving struggling readers' reading skills during Year 2 for FSM pupils?

A compliance analysis using a complier average causal effect (CACE) approach was also conducted based on dosage, that is, the amount of time pupils spent using the online activities.

The impact evaluation was designed to answer the primary and secondary research questions as described above. Alongside this an implementation and process evaluation (IPE) was conducted. The IPE was designed to be complementary to the impact evaluation and exploratory in terms of the impact outcomes. It sought to answer the following research questions:

1. What is the relationship between the fidelity of the intervention and the impact on pupil outcomes?
2. How much variability occurred in implementation of the intervention across different settings in respect of:
 - a) dosage - the number and frequency of overall sessions?
 - b) school factors, such as physical space/place of intervention?
 - c) teacher factors, including supervision of the sessions and use of paper-based resources?
 - d) potential or actual perceived barriers to implementation?
3. What was the reach of the intervention (including the use of the training and support provided)?
4. What was the nature of teaching as usual (the control conditions)? What support was offered to those pupils not allocated to the intervention group.

The evaluation protocol can be found at:

[https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Protocols/Lexia_protocol_\(amended\).pdf](https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Protocols/Lexia_protocol_(amended).pdf)

The Statistical Analysis Plan (SAP) can be found at:

https://educationendowmentfoundation.org.uk/public/files/Projects/Lexia_SAP.pdf

Ethics and trial registration

Ethical approval for this study was granted by the Education Ethics Committee, University of York on 13 December 2017, reference 17/26.

All settings that took part in the trial completed a MOU during the recruitment phase. See Data Protection section below for more details.

The trial has been registered at the International Standard Randomised Controlled Trial Number (ISRCTN) registry: ISRCTN42120944, <https://doi.org/10.1186/ISRCTN42120944>.

Data protection

All personal data was processed under Article 6 (1) (e), 'processing necessary for the performance of a task carried out in the public interest', of the General Data Protection Regulation (GDPR) and the Data Protection Act 2018. The University of York was deemed to be a data controller (as defined by the data protection legislation) with regard to the personal data used for this project.

Schools were informed of the data requirements through the MOU (Appendix 1) and MOU Addendum (Appendix 2). A DSA (Appendix 3) and DSA Addendum (Appendix 4) was put in place between the University of York and each school which included full details of the types of personal data being shared, the purpose and duration of that sharing, and the responsibilities each party had in relation to that information.

After teachers identified the pupils in the lower half of their class in terms of reading ability, the parent or carers of those pupils received a Parent/Carer Information Sheet (PIS) (Appendix 5) that outlined the information the evaluators would collect and how it would be used. It incorporated a Participant Withdrawal from Research Form, which gave parent and carers the option of withdrawing their child's data from data collection and processing.

All electronic data transfer, to and from settings, was via encrypted spreadsheets sent through the University of York's secure DropOff service. The trial management system and trial data were held separately on secure University of York servers with access limited to specified members of the evaluation team.

All results have been anonymised so that no school or individual student will be identifiable in the report or dissemination of any results.

All data held by the evaluation team will be retained for three years after publication of the final report and then securely destroyed.

For the purpose of research, the pupil data will be linked with information about the pupils from the National Pupil Database (NPD) and shared with the DfE, the EEF's archive manager and, in a pseudonymised form, the Office for National Statistics and potentially other research teams. Further matching to NPD data may take place during subsequent research. Once the data is transferred to the EEF, the EEF becomes a data controller.

Project team

The independent evaluation team was led by Dr Louise Tracey and Louise Elliott (co-principal investigators) who were responsible for the day-to-day management and coordination of the evaluation. Caroline Fairhurst, as co-investigator, led on the randomisation and oversaw the statistical analysis of the study. Laura Mandefield was the project statistician who undertook the analysis and Sarah Ellison was the research associate. Imogen Fountain was the project support officer and was responsible for liaising with schools, uploading and monitoring the teacher surveys, scheduling the assessments and the school observation visits, and maintaining the project database. Overall, the evaluation team was responsible for the design, randomisation, data collection, analysis, and reporting of the evaluation. The members of the evaluation team were based in the Department of Education and York Trials Unit at the University of York.

The delivery team was based in the School of Social Sciences, Education and Social Work at Queen’s University, Belfast. The team was led by Professor Allen Thurston. Dr Maria Cockerill was responsible for school recruitment and liaising between stakeholders to monitor progress of the trial. Dr Joanne O’Keeffe was the research associate.

Both the evaluation and delivery teams worked closely with LexiaUK, which provided the licences, training, and support to schools during the trial. This work was led by Rob Kay, Director of LexiaUK, assisted by Norman Robinson, Chief Operating Officer.

Methods

Trial design

The evaluation was an efficacy RCT with randomisation at individual pupil level within schools (Table 4). This provided the ideal counterfactual, avoiding the issue of variation between schools, and controlling for selection bias through randomisation. As the intervention is delivered one to one via a computer, the possibility of diffusion is reduced. Although the logic model posits school change this would be limited for those children not involved in the Lexia programme and would impact control and intervention children similarly, so it was felt to be a strong design. Our sample comprises struggling readers within Year 2 in evaluation schools in the academic year 2018/2019.

Table 4: Trial design

Trial design, including number of arms		Two-armed efficacy trial.
Unit of randomisation		Pupil.
Stratification variable		School.
Primary outcome	Variable	Reading ability.
	Measure (instrument, scale, source)	WRMT-III composite of the scores of four subtests—Word Identification, Word Attack, Passage Comprehension, and Oral Reading Fluency. Range 231–580. Pearson Assessment.
Secondary outcome	Variable	Word recognition.
	Measure (instrument, scale, source)	WRMT-III Word Identification subtest score. Range 55–145.* Pearson Assessment.
Secondary outcome	Variable	Decoding.
	Measure (instrument, scale, source)	WRMT-III Word Attack subtest score. Range 64–145.* Pearson Assessment.
Secondary outcome	Variable	Comprehension.
	Measure (instrument, scale, source)	WRMT-III Passage Comprehension subtest score. Range 57–145.* Pearson Assessment.
Secondary outcome	Variable	Fluency.

	Measure (instrument, scale, source)	WRMT-III Oral Reading Fluency subtest score. Range 55–145.* Pearson Assessment.
Secondary outcome	Variable	Reading attainment.
	Measure (instrument, scale, source)	KS1 national test reading raw score. Range 0–40. Obtained from schools.
Baseline for primary outcome	Variable	Reading Ability, Fluency, and Reading Attainment.
	Measure (instrument, scale, source)	WRMT-R/NU composite of the scores of three subtests —Word Identification, Word Attack, and Passage Comprehension. Range 74–166. Pearson Assessment.
Baseline for secondary outcome	Variable	Word Recognition.
	Measure (instrument, scale, source)	WRMT-R/NU Word Identification score. Range 70–159. Pearson Assessment.
Baseline for secondary outcome	Variable	Decoding.
	Measure (instrument, scale, source)	WRMT-R/NU Word Attack score. Range 81–163. Pearson Assessment.
Baseline for secondary outcome	Variable	Comprehension.
	Measure (instrument, scale, source)	WRMT-R/NU Passage Comprehension score. Range 70–175. Pearson Assessment.

* Lower bound differs depending on age.

Changes to the original trial design were required. The criterion for pupil eligibility changed because fewer children than anticipated were fulfilling the original selection threshold.⁴ The number of pupils recruited per school was increased from nine to a maximum of 14 because schools suggested that on average they could manage up to seven Lexia pupils (therefore, double this number could be recruited per school and randomised 1:1 to intervention or control).

Those pupils not allocated to the Lexia intervention continued with teaching as usual during the intervention period; however, participating schools received a two-year licence for Lexia Reading Core5 for use with 30 pupils. The licences could be used for additional pupils, outside of Year 2, during the study period and with any pupils within the school (including Year 2) for the remainder of the licence to bring the total usage to 30 pupils.

No financial incentive was provided but the programme was provided at a reduced cost of £500 plus VAT per school for a two-year licence. This fee was non-refundable however, the licence was contingent on schools delivering Lexia Reading Core5 to the Year 2 pupils selected to receive it during the study period and would have been terminated should a school withdraw from the evaluation during the period of the study. This also meant that control group pupils could receive the intervention in the following year if the school deemed it necessary.

Participant selection

The delivery team at Queen's University Belfast were responsible for recruiting schools and initially concentrated recruitment in the North East and Yorkshire and Humber regions at the EEF's request as part of their North East Literacy

⁴ Amended protocol:

[https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Protocols/Lexia_protocol_\(amended\).pdf](https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Protocols/Lexia_protocol_(amended).pdf)

Campaign.⁵ Most of the recruitment was undertaken by emailing and telephoning contacts in these areas with specific focus on schools with above average 'everFSM' pupils.⁶ Other geographical areas were then included to ensure the recruitment target was met. The areas were Wiltshire, Lancashire, and Greater London. Schools from these areas were accepted to the trial in groups providing there was a sufficient number to enable cost-effective data collection. There were a number of criteria that the schools had to meet in order to be eligible to participate:

- having approximately 50 pupils per year group;
- not being involved in any other EEF trial focusing on KS1 literacy or aiming to achieve change at a whole school level;
- not already using Lexia Reading Core5 or not having used it in the 12 months prior to the recruitment phase;
- being able to provide the necessary IT technology required in order to support the intervention; and
- being willing to implement the intervention with respect to the random allocation (that is, only with those pupils assigned to the intervention group).

Schools were approached and those that responded were sent an information sheet and an Expression of Interest (EOI) form to complete. Schools that returned the EOI were screened for eligibility and eligible schools were sent a MOU to sign and return. The MOU included permission for the evaluation team to access school data generated by the Lexia Reading Core5 software. Upon receipt of the signed MOU, specified data requested in the MOU (including pupil UPNs, pupils' FSM status, and KS1 reading raw and scaled scores) and contact details for their headteacher, lead contact, and class teachers were shared by the school with the evaluation team.

All recruited schools were sent a DSA and, as recruitment took part in early 2018, they were later sent an MOU Addendum to comply with the new GDPR, which came into effect in May 2018. Schools had to accept and sign the DSA to continue participating in the evaluation.

The pupils eligible to take part in the trial were the lower half of the Year 1 pupils in terms of reading ability as identified by the school. The school was sent a PIS by the evaluation team to distribute to the parents of the identified pupils. The PIS gave parents the option to withdraw their child's data from the study (as randomisation was at pupil level, with a limited number of participants, this effectively meant withdrawing them from the evaluation as a whole and thus the possibility of receiving the intervention). Explicit consent was not sought from parents as the data was being processed under Article 6 (1) (e) of the GDPR: 'processing is necessary for the performance of a task carried out in the public interest'.

However, a small number of schools were unwilling to distribute the PIS without requesting explicit consent from parents and carers. This was accepted on the understanding that the PIS was unchanged apart from the withdrawal statements being replaced by explicit consent and that the evaluation team approved the document before circulation.

The schools were then asked to prepare a spreadsheet with the details of the pupils eligible to participate in the trial, that is, those that had been:

- identified as being in the lower half of the year group in terms of reading ability; and
- a withdrawal of data form had not been received from the parent/carers.

The schools provided the evaluation team with an encrypted file securely using the University of York's DropOff service. Full guidance for doing this was emailed to each school.

The children were independently assessed by the administration of the Word Identification, Word Attack, and Passage Comprehension subtests of the Woodcock Reading Mastery Tests, Revised Normative Update (WRMT-R/NU), by assessors recruited and trained by the University of York. Originally, it was planned that pupils who received a standard age score (SAS) of 85 or less (classed as below average: Castro and Robnolt, 2016) in any of the three subtests of the WRMT-R/NU at pre-test would be eligible to participate in the study. This was amended after pre-testing because fewer children than anticipated fulfilled the original selection criteria.

⁵ https://educationendowmentfoundation.org.uk/scaling-up-evidence/campaigns/north-east-literacy-campaign/?utm_source=site&utm_medium=search&utm_campaign=site_search&search_term=North%20east

⁶ Those identified in the NPD as having been FSM-eligible in the previous six years.

Schools indicated that they were able to deliver Lexia to six or seven pupils, a higher number than expected; therefore, it was decided to include the 12 to 14 pupils in each class with the lowest pre-test score. If there was more than one pupil with the same score as the 12th pupil then up to 14 pupils were included. If more than 14 pupils would have been included by having the same score as the 12th ranking pupil then we planned to randomly select those with this score for inclusion such that a maximum of 14 were included per school so that there was no potential for more than seven pupils to be randomised (1:1) to the Lexia intervention. However, in the event, this scenario did not arise. The implications of using a relative cut-off were that some pupils in some schools were not strictly struggling readers but we had to balance this against having fewer pupils. It was felt important that eligibility was determined independent of the teacher judgement to ensure consistency across participating schools.

Schools were eligible for randomisation after:

- signing the MOU;
- providing specified data as requested in the MOU and contact details for their headteacher, lead contact, and class teachers;
- all baseline testing—of pupils identified by the class teacher as being in the lower-half of the class in terms of reading—had been completed; and
- all teachers in Year 2 had completed an online pre-randomisation survey.

Outcome measures

Baseline measures

The WRMT-R/NU was administered one to one to all the participating pupils by trained assessment administrators before randomisation at the end of the academic year 2017/2018. Standard age scores from this were used as the baseline measure of prior attainment (Table 5).

The WRMT-R/NU is the previous version of the Woodcock Reading Mastery Test (its content was developed in 1987), which was superseded by the third version (WRMT-III) in 2011. The initial intention was to use the WRMT-III as the pre-test, however, upon ordering the tests the evaluation team were informed that they could not be delivered in time for the testing period and so an older version of the test had to be administered (the evaluation team already had copies of this test from a previous research project). This was not deemed to have a negative impact, however, given that all children received the same assessment and therefore scores would be comparable.

All the subtests were fully revised and updated between the NU version and the III version of the test instrument. The III version used current research to ensure it covered all areas of reading ability and it was re-normed and can be used from an earlier age (four years and six months rather than five years). The newer version contains more engaging content, has two equivalent test forms, and the item content was updated to ensure no bias (relating to gender, ethnicity, socioeconomic background, and so forth). For the subtests we were using, there are fewer items on each test (about half as many) and about 40% of the items are new for Word Identification and Word Attack, while for Passage Comprehension all but one items are new. Now the rule is to discontinue the test if four consecutive items are incorrect rather than six consecutive items at the end of a page with the NU version. The WRMT-R/NU does not include a fluency subtest like the WRMT-III so the composite score is used as the measure of prior attainment for the fluency subtest of the post-test (Table 5). The composite standard age score was calculated where a pupil had a valid score for at least two out of the three subtests.

Table 5: Outcome measures and associated baseline measures of prior attainment (all scores are age-normed)

Measure	Variable	Scoring range	Measure	Variable	Scoring range

OUTCOME (measured June 2019)			BASELINE (measured June/July 2018)		
WRMT-III composite reading score*	Reading ability	231–580	WRMT-R/NU composite reading score**	Reading ability	74–166
WRMT-III word identification score	Word recognition	55–145 [^]	WRMT-R/NU word identification score	Word recognition	70–159
WRMT-III word attack score	Decoding	64–145 [^]	WRMT-R/NU word attack score	Decoding	81–163
WRMT-III passage comprehension score	Comprehension	57–145 [^]	WRMT-R/NU passage comprehension score	Comprehension	70–175
WRMT-III oral reading fluency score	Fluency	55–145 [^]	WRMT-R/NU composite reading score***	Reading ability	74–166
KS1 raw reading score	Reading attainment	0–40	WRMT-R/NU composite reading score	Reading ability	74–166

* Primary outcome.

** The average SAS score calculated at baseline and used to identify pupils for the evaluation.

*** No fluency subtest in pre-test.

[^] Lower bound differs depending on age.

Primary outcome

The primary outcome measure was a composite of the standard age scores of four subtests of the WRMT-III—Word Identification, Word Attack, Passage Comprehension, and Oral Reading Fluency (Woodcock, 2011)—for Year 2 pupils. The composite score was calculated where a pupil had a valid score on all four of the subtests. The WRMT-III is a standardised measure suitable for ages four years and six months to 79 years 11 months. It is paper-based and it took approximately 20 to 25 minutes in total to administer the four subtests to each child. The composite score constructed from the four subtests is designed to reflect overall reading ability. In the protocol, we state that raw scores would be used for the post-test WRMT-III: by this we meant fine graded scores, although in the event we have used the standard age scores, following guidance from the user manual, as the composite score is obtained by first calculating standard age scores for each subtest and second by summing these scores to produce a score between 231 and 580. A higher score indicates more proficient reading ability. ‘Raw’ scores have no meaning in this instrument.

This composite was considered an appropriate measure as the four subtests identify the key areas in which readers typically struggle and those that Lexia Reading Core5 targets. We considered adding the Word Comprehension subtest to enable us to use the Total Reading Cluster as an overall measure. However, in consultation with the EEF it was decided not to do this as Word Comprehension would replicate some of the skills involved in Passage Comprehension and would require additional time to conduct, which would have resulted in more burden on children and schools and increased costs.

Administration of the WRMT-III at post-test was conducted one to one with each child by trained administrators who were blind to group allocation to avoid the potential for ascertainment bias.

Secondary outcomes

The secondary outcome measures are the standard age scores of the individual subtests of the post-test WRMT-III and the KS1 reading raw scores. The KS1 reading test consisted of two papers administered nationally to pupils by schools

during May 2018.⁷ The KS1 reading scores were securely transferred from schools in an encrypted Excel spreadsheet for the Year 2 pupils as the NPD only holds data on whether pupils are 'working towards', 'working at', or 'working above' the standard expected at the end of KS1. Word Identification and Oral Reading Fluency scores have a range of 55 and 145, Word Attack scores have a range between 64 and 145, and Passage Comprehension scores range between 57 and 145. KS1 raw reading scores are scored between 0 and 40. In all secondary outcomes, a higher score indicates a better outcome.

Sample size

Table 8 presents a summary of the sample size assumptions and MDES at protocol, randomisation, and analysis stages.

Protocol

We proposed to recruit 57 schools; this number of schools was considered realistic given the capacity of the delivery team. Assuming an average of 56 pupils in the school (28 per class, two-form entry), we estimated that an average of nine pupils per school would be identified as struggling readers. This was based on administration of a similar assessment, the York Assessment for Reading Comprehension (YARC), with similar pupils in a previous study (Tracey et al. 2014), which found that 17% of pupils were struggling readers. Assuming a pre- and post-test correlation of 0.6 between the baseline and post-test of the WRMT, with 80% power, the minimum detectable effect size (MDES) with this sample size would be 0.20 allowing for 10% pupil-level attrition (StataCorp., 2017). The assumed correlation of 0.6 is based on the test-retest reliability of the WRMT-III of 0.51 to 0.95,⁸ and an observed pre-post test correlation of 0.64 for the YARC Comprehension subscale, a similar reading comprehension test administered in a similar population to this trial (personal correspondence with Louise Elliott for the Success for All trial: Tracey et al., 2014). The model that was used to analyse the data was adjusted for any potential within-school correlation but this was not incorporated in the sample size calculation.

Though the trial was not powered to detect an effect in the FSM subgroup, an analysis in the subset of FSM pupils was proposed as a secondary investigation. The first 50 schools recruited to the trial had an average FSM of 29.3% overall. Based on this estimate, we might conservatively have expected three of the struggling readers per school to be eligible for FSM so in this subgroup—assuming other parameters are equal—the MDES would be 0.36. However, it is likely that FSM status and being a struggling reader are correlated, so a higher proportion of the nine identified struggling readers might have been eligible for FSM. For example, with an average of six FSM pupils per school, the MDES would be approximately 0.26 (all other things being equal).

Randomisation

A total of 57 schools were recruited. All eligible pupils with pre-test WRMT-R/NU scores of less than or equal to the 12th lowest ranking pupil in the class were randomised (1:1) to either the Lexia intervention or teaching as usual; 697 pupils in total (mean 12.2 pupils per school, range 12 to 13 pupils) were randomised within school. Assuming a pre- and post-test correlation of 0.6 and 10% pupil-level attrition, the MDES with this sample size would be 0.18 with 80% power for the overall study.

For the FSM subgroup, since the recruited schools had an average of 29.2% FSM pupils, we could conservatively estimate an average of four randomised pupils to be eligible for FSM in each recruited school. Using otherwise identical parameters to those described above, the MDES in this subgroup would be 0.31 with 80% power. However, there could be more than four randomised FSM pupils in each school as FSM status is correlated to pre-test scores. Depending on the percentage of FSM pupils among the randomised pupils, MDES with 80% power could be as low as 0.18.

Analysis

⁷ <https://www.gov.uk/government/publications/key-stage-1-tests-2018-english-reading-test-materials>

⁸ <https://docplayer.net/45893790-Wrmt-iii-anise-flowers-ph-d-september-copyright-pearson-education-and-its-affiliates-all-rights-reserved.html>

Most of the assumptions in the original sample size calculation held; however, the pre- and post-test correlation was found to be 0.75, which was higher than our original estimate of 0.6. The average cluster size in the analysis was 10.9. Although not included in our original sample size, an ICC at the school level was estimated to be 0.06. Including this further adjustment for clustering at school level allowed an MDES of 0.19 at analysis.

FSM status was obtained from the NPD. Among the FSM subgroup an average cluster size of 4.16 was observed. The ICC was 0.075 and correlation between pre- and post-test scores was 0.74. Therefore, with 80% power and two-sided $\alpha=0.05$, the estimated MDES for this subgroup is 0.29.

Randomisation

Randomisation was completed before the end of the summer term 2018. Block randomisation, stratified by school, was used to randomly allocate pupils within schools to either the intervention or control group. The original protocol indicated that variable block sizes would be used, but a fixed block size of two was implemented in the end to ensure no more than seven pupils were allocated to receive the Lexia intervention in any one school. The allocation schedule was generated by the trial statistician, Caroline Fairhurst, in Stata v15. The allocation schedule was then populated by a different member of the evaluation team who ordered eligible pupils within each school according to their pre-test score and copied the pupil IDs across to the schedule to obtain their allocation, thereby ensuring allocation concealment from the schools with no chance of subversion.

The statistician was not blind to allocation for analysis but in order to minimise the risk of bias, the statistical analysis plan was prespecified and signed off before the statistician, Laura Mandefield, received the final analysis dataset.

Statistical analysis

The statistical analysis followed the most recent revised EEF Statistical Analysis Guidance (2018) and the statistical analysis plan agreed for the study (Fairhurst and Rex, 2019). Analysis was conducted in R v3.5.3 (R Development Core Team and R Core Team, 2011) using the principle of intention to treat (ITT), where data was available, including all pupils in the groups to which they were randomised irrespective of whether or not they actually received the intervention.

Statistical significance was assessed using two-sided tests at the 5% level. Estimates of effect with 95% confidence intervals (CIs) and p-values have been provided.

Pre- and post-test scores have been summarised by trial arm and are presented using histograms. Characteristics of the recruited schools were presented overall. Pupil characteristics (age, gender and FSM) and measures of prior attainment were summarised descriptively by randomised group both as randomised and as analysed in the primary analysis. Gender and FSM status were sourced from the NPD. Continuous measures are reported as a mean and standard deviation (SD) while categorical data is reported as a count and percentage. The unadjusted difference between groups on the pre-tests were reported as a Hedges' g effect size with 95% CI.

Primary analysis

The primary analysis assessed any difference in the WRMT-III composite reading score between the two groups. The primary analysis was conducted using a linear mixed model at the pupil level with outcome score as the response variable. Group allocation and measure of prior attainment were included as fixed effects in the model, and school as a random effect. The model equation was as follows:

$$Y_{ij} = \beta_0 + u_{0j} + \beta_1 \text{Prior}_i + \beta_2 \text{Rand}_i + e_{ij}$$

Y_{ij} = response of the i^{th} pupil in j^{th} school, $i=1, \dots, n$, $j=1, \dots, m$

n = number of participating pupils

m = number of recruited schools

Prior_i = score for measure of prior attainment for the i^{th} pupil

$Rand_i$ = indicator variable for group allocation for i^{th} pupil (0 = control, 1 = intervention)

u_{0j} = random effect represents the deviation of school j 's mean z-score from the grand mean

$$u_{0j} \sim N(0, \sigma_2^2)$$

e_{ij} = residual error term for the i^{th} pupil in j^{th} school

$$e_{ij} \sim N(0, \sigma_1^2)$$

The normality of the standardised residuals was checked using a histogram and QQ plot and the homoscedasticity of the residuals assessed using a scatter plot of fitted values against the residuals. Visual inspection of the plots only was used (no formal statistical tests).

Secondary analysis

The secondary outcomes of the individual WRMT-III subtest standard age scores and KS1 reading raw score were analysed as described for the primary outcome using an appropriate pre-test WRMT-R/NU score as the measure of prior attainment in the model: for example, WRMT-R/NU Word Identification subtest score at baseline was used as the measure for prior attainment for the analysis of the WRMT-III Word Identification post-test outcomes. The composite WRMT-R/NU pre-test score was included in the model for the outcome of KS1 reading.

Analysis in the presence of non-compliance

The programme was delivered from September 2018 over a period of 24 weeks—two full terms. Schools were expected to schedule four sessions of 30 minutes (including ten minutes set-up time) per week.

Some pupils received more than the initially intended number of sessions as schools were allowed to continue using the online activities with the intervention group pupils beyond the initial 24 week delivery period prior to post-testing providing these licences continued to be used for the Year 2 intervention pupils only (the use of the remaining licences continued as previously specified). Data on individual participant use of Lexia was exported from the system and included the date and duration of each Lexia session at the pupil level. The number and length of sessions received by pupils are summarised in the section on Outcomes Analysis below.

A Complier Average Causal Effect (CACE) analysis for the primary outcomes was used to account for compliance with the intervention. Three analyses were conducted at the pupil level, considering compliance as:

- a dichotomous variable (complied or not) with compliance defined as completing at least 60 minutes a week (excluding set-up time) for a minimum of 12 (not necessarily consecutive) weeks;
- a dichotomous variable (complied or not) with compliance defined as completing at least 720 minutes in total; and
- a continuous measure of number of hours of the intervention received.

These definitions were derived through discussion with the Lexia delivery team during the IDEA meeting. The first two constitute an approximate 50% of the full possible exposure, which was deemed the minimal acceptable threshold; however, the first definition implies that the sessions are delivered in a more structured manner to the second definition. Pupils who satisfy definition one will automatically satisfy definition two, but not vice versa. A two-stage instrumental variable approach was used with random group allocation as the instrumental variable (Dunn, Maracy and Tomenson, 2005).

Missing data analysis

The amount of missing baseline and outcome data has been summarised, and reasons for missing data explored and provided, where available. Significant predictors and possible mechanisms for the missing data are discussed (see Outcomes Analysis section below). A mixed effect logistic regression model was run to predict the presence of missing primary outcome (composite WRMT-III reading score) data and assess the missing data mechanism including group allocation and WRMT-R/NU pre-test score, with school as a random effect. The impact of missing data on the primary

analyses was additionally assessed using multiple imputation by chained equations, including all available baseline variables (pre-test WRMT-R/NU scores, age, gender and FSM status⁹; Azur et al., 2011).

A 'burn-in' of 150 was used and 30 imputed datasets were created. The primary analysis was then rerun within the imputed datasets and Rubin's rules were used to combine the multiply-imputed estimates. This multiple imputation model, however, did not account for the multilevel nature of the data; therefore, additional sensitivity analyses were carried out by imputing the missing data using multilevel imputation using the 'jomo' package in R (Quartagno and Carpenter, 2018) and repeating the primary analysis with this imputed dataset. The result from this was compared against the result from the data imputed using MI by chained equations for major differences.

Subgroup analyses

The effect of the intervention on pupils who were eligible for FSM was assessed via the inclusion of FSM status (using the EverFSM indicator EVERFSM_6_P from the NPD) and an interaction term between FSM status and allocation in the primary analysis model. Additionally, the primary analysis was repeated in the subgroup of pupils who were ever eligible for FSM.

Estimation of effect sizes

The original protocol stated that the treatment effect size would be calculated based on the adjusted mean difference between the intervention and control group and the total variance (between plus within school variance) obtained from the linear mixed model. However, this did not follow the most recent EEF analysis guidance (EEF, 2018) so the protocol was amended to reflect the new guidance. Therefore, effect sizes were calculated by dividing the adjusted mean difference between the intervention and control group by the pooled variance obtained from the unconditional model. A 95% CI for the effect size was calculated by dividing the 95% confidence limits for the adjusted mean difference by this same variance. All parameters used in these calculations have been provided in the final report.

Effect sizes have also been presented in terms of estimated additional months of progress based on EEF guidelines,¹⁰ for example, an effect size in the range -0.04 to 0.04 translates to zero months' additional progress, 0.05 to 0.09 as one month, 0.10 to 0.18 as two months, and so on.

Estimation of ICC

The intracluster correlation coefficient (ICC) associated with school for the observed pre- and post-test outcomes are presented with 95% CIs calculated using bootstrapping methods.

Implementation and process evaluation (IPE)

Research methods

The IPE research questions, as indicated above, were addressed via a pragmatic, mixed methods approach including the following elements:

Pre and post teacher survey

All teachers in Year 2 were asked to complete an online pre-randomisation survey (spring/summer 2018) to establish a full picture of school and teacher contextual factors: literacy teaching in Year 2, what usual or baseline practice for struggling readers looks like in the participating schools, and IT facilities within schools—including support, familiarity, and experience relating to IT (n = 135). A post-test survey was administered to the Year 2 teachers and teaching assistants involved in delivering the programme in all 57 schools (summer 2019, n = 98). This survey contained questions relating to use of the online student activities (for example, when scheduled and whether delivered in a group timeslot or individual timeslots, what activities other Year 2 pupils are engaged in at that time, and the training and

⁹ Version 1.0 of the report, published in September 2021, included multiple imputation models without the presence of gender and FSM status as predictors due to delays accessing NPD data via the ONS Secure Research Service during the COVID-19 pandemic. These results have now been moved to Appendix E in this updated version of the report.

¹⁰ <https://educationendowmentfoundation.org.uk/help/projects/the-eefs-months-progress-measure/#:~:text=For%20EEF%20trials%2C%20the%20Months,who%20did%20not%20receive%20it>

support used) and associated benefits (for example, 'softer' outcomes in terms of pupils' confidence and engagement) and challenges (for example, IT support, scheduling time in the school day). The two surveys (pre-test and post-test) were designed after discussion with the Delivery Team, administered online using Qualtrics software and piloted with a similar cohort in advance.¹¹

School visits

Ten schools were visited to form case studies for a more in-depth process evaluation. They were chosen on a pragmatic basis to allow for cost-efficient use of resources but with a spread of FSM across the schools and geographically. All ten schools contacted to participate in the case study visits agreed to participate in the IPE. There were three elements to the school visits, as described below.

- Observation of Lexia implementation

Visits were scheduled to coincide with a time when Lexia Reading Core5 was being delivered to observe implementation of the online element, including what space and IT was provided, support offered to children during the Lexia Reading Core5 sessions (for example, teacher or teaching assistant present), number of children in any one session, and pupil engagement. In total, ten sessions were observed (one in each school) lasting approximately 30 minutes each. The observation schedule was designed based on items of interest for the IPE and the programme logic model—and after a demonstration of the online elements by the delivery team for the benefit of the evaluation team. It was originally planned to observe control pupils within these schools, however, given that any supplementary activities were taking place at the same time as the intervention, this aspect was dropped from the evaluation for pragmatic reasons and covered in the interviews instead.

- Teacher interviews

Year 2 teachers were interviewed to establish school and class contextual practices—the timing and space provided for delivery, additional support for children using the programme, provision for those struggling readers allocated to the control condition, and the rationale behind these decisions. In addition, there was discussion about whole-class literacy provision for Year 2. Information was also elicited regarding the ease of use of Lexia Reading Core5, the training and support used, and attitudes towards the programme. If teaching assistants were involved in programme delivery and were available during the visit (outside of the observation of online student activities), they were interviewed regarding their experience of delivering the programme. Sometimes the Lexia lead (not a Year 2 teacher or TA) also took part in the interview. Eleven interviews took place (in one school the TA was interviewed separately) involving eight TAs, nine teachers, two Lexia project leads (not a Year 2 teacher or TA), and one assistant headteacher. Although there could be a conflict of interest by interviewing TAs and teachers together on no occasion did the researchers feel that this was an issue in terms of the reliability of the data collected.

- Pupil focus groups

We conducted focus groups with Year 2 struggling readers allocated to the intervention condition in five out of the ten schools visited. Schools were selected based on the number of children on EVERFSM (in order to provide a range from those above and below average) and a range of geographical locations. Children for the focus groups were selected by the Year 2 teacher based on parental consent to participate and school judgement. Researchers also checked at the beginning of each focus group that the child was happy to participate. Whilst we appreciate this was not a random selection it was deemed to be the most ethically appropriate approach given the age of the children. The focus group was designed to be framed within our observation of Lexia sessions in order to allow researchers to assess pupil experience of using the online student activities, their perceptions of their learning, pupil engagement, and confidence. These focus groups usually took place directly after the Lexia session to aid recall and focus and the researcher was able to use prompts based on their observation of the session in order to facilitate a meaningful dialogue between children and researchers. Each focus group involved three or four pupils from the intervention group.

¹¹ <https://www.qualtrics.com>

All school visits were conducted by the two principal investigators. The first visit was conducted jointly to check that the instruments worked effectively in schools and could be adapted if necessary (in the event, no changes occurred) and to ensure shared understanding of the items and processes for data collection and, in the case of the observations, to check for inter-rater reliability. There was approximately 90% agreement between researchers. One researcher conducted six visits and the other five visits (including the joint visit—ten visits in total).

Pupil usage data

The evaluation team was provided with access to the data files produced by, and for, schools via the software following opt-in permission obtained from schools in the MOU and then confirmed by the DSA. Only one school chose not to participate in this aspect of the evaluation. The Lexia Reading Core5 software provided school-, class-, and individual-level reports, which were used to gather information relating to implementation, fidelity, and dosage and are routinely used by LexiaUK and schools implementing the programme. This information was particularly valuable for both the impact and process evaluations as it reduced the potential burden of the research on schools (see CACE analysis section). The data provided information about how often the online student activities were used and for what length of time (that is, frequency and length of time the child was logged into the Lexia ILS software), any patterns of usage during the programme delivery period, and what areas of the adaptive software showed the greatest amounts of progress over time. There was no data available on use of paper-based activities via the software although evidence on this was collected via the surveys and interviews (as described above).

Analysis

Data on pupil usage of the online activities was downloaded into SPSS 26 (IBM, 2019) and analysed primarily using descriptive analysis techniques. Surveys were downloaded into Excel, and analysed at school level or individual level depending on questions asked. Free text responses were also analysed in this way providing both quantitative and qualitative data. All analysis focused on the research questions although further, related analysis occurred to ensure that important information was not missed.

Interview and focus group data was transcribed and imported into NVivo 12 software (QSR, 2018). It was analysed thematically, based on the relevant research questions, using a deductive approach. Finally, the interviews were triangulated with the observation and pupil focus group data (where pupils focus groups occurred). The use of a case study approach was posited in the protocol but after agreement it was felt that, given the small number of schools visited and the even smaller number of pupil focus groups, to do so would not be sufficiently pseudonymised. Where links have been able to be made they have, however, been done so, but not identified within the text. Consequently, in some cases results have been reported at the school level, which has meant that survey responses have been collated and reported together (that is, teacher and TA survey responses). In addition, some interviews took place with both the teacher and the TA at the same time and there were felt to be no differences in reporting when compared to interviews occurring separately. We did not, however, feel it appropriate to include the pupil focus groups in this. If there were any disparities in responses these were intended to be highlighted in the text but in the event these did not occur. Consequently, we feel secure in the findings.

Table 6 provides a summary of the IPE data collected, the form of analysis that was conducted, and its relationship to the research questions and logic model.

Table 6: IPE methods overview

Research methods	Data collection methods	Participants/data sources	Data analysis methods	IPE Research questions addressed	Implementation/logic model relevance
Pupil usage data	Pupil usage data	Generated by Lexia software and provided by LexiaUK	Statistical analysis in SPSS	1. 2.a, 2.b, 2.e	Dosage, compliance

Survey	Teacher survey	Pre-test all Year 2 teachers in participating schools Post-test Year 2 teachers and TAs	Statistical analysis in Excel	1. 2.a, 2.b, 2.c, 2.f 3. 4.	Usual practice, implementation
Observation	Observations	Research observation of Yr 2 Lexia session	Hand analysis	1. 2.b, 2.c, 2.e	Implementation
Interview	Interviews	Yr 2 teachers/teaching assistants	Deductive, thematic analysis in NVivo	1. 2.a, 2.b, 2.c, 2.e, 2.f 3. 4.	Implementation, compliance, dosage, usual practice
Focus group	Pupil focus groups	Year 2 intervention group pupils	Deductive, thematic analysis in NVivo	1.	Compliance

Costs

The researchers sought to assess the initial cost of implementing the programme and the cost per pupil of implementing over a three-year period following EEF guidance (EEF, 2019). LexiaUK offer multi-year packages but for this trial the EEF agreed a two-year package and this has been taken into account in the costing.

The cost of the intervention, including software licences, ongoing support, any prerequisite resources, and time required for staff training (headteacher, teachers, and teaching assistants) was collected from the developers. Costs relating to any additional materials or resources needed within schools for the intervention was collected through the teacher survey at the end of the intervention period. Additional questions were asked during the teacher interviews to identify any issues around provision and the cost of resources that may not have been previously identified.

The survey was also used to collect information about the teacher and teaching assistant time required to facilitate the intervention, including time taken delivering the intervention, training time (other than the initial training, collected from the developer), and time spent on any other activities related to the intervention.

Timeline

Table 7 highlights the main activities undertaken during the course of the evaluation of the Lexia Reading Core5 intervention.

Table 7: Timeline

Dates	Activity	Staff responsible / leading
June 2017–October 2017	Set Up Meetings 1 and 2 and IDEAS meeting	Evaluation Team Delivery Team / QUB, LexiaUK
December 2017	Ethical approval granted from University of York Education Ethics Committee	Evaluation Team
January–June 2018	Recruitment of schools including MOUs signed and returned and passed to evaluation team	Delivery Team / QUB
April–June 2018	Licences for Lexia issued to schools	Delivery Team / LexiaUK
April–June and September 2018	Training for Lexia deliverers in schools	Delivery Team / LexiaUK
May–June 2018	Teacher baseline survey data collection	Evaluation Team
May–July 2018	Pupil data collection and validation	Evaluation Team

June–September 2018	DSA and MOU Addendum collection	Evaluation Team
June–July 2018	Baseline (pre-test) assessment (WRMT/NU) visits to schools	Evaluation Team
July 2018	Input and validation of pupil pre-tests	Evaluation Team
July 2018	Randomisation of pupils	Evaluation Team
August 2018	Pupils randomised to receive Lexia set up on each school account	Delivery Team / LexiaUK
September 2018	Intervention starts	Delivery Team / LexiaUK and schools
January–February 2019	IPE School observation visits, interviews and focus groups	Evaluation Team
March 2019	ISRCTN registration assigned	Evaluation Team
May–July 2019	Post-test assessment (WRMT-III) visits to schools	Evaluation Team
June–July 2019	Teacher and TA post-test survey data collection	Evaluation Team
July 2019	Input and validation of pupil post-tests	Evaluation Team
September 2019	NPD application	Evaluation Team
October 2019	Analysis of pre and post-test data	Evaluation Team
December 2019 ^a	Draft report due	Evaluation Team
December 2019	Data collection of KS1 results	Evaluation Team
May 2021	Draft report submitted	Evaluation Team
October 2021	Accessed NPD Data	Evaluation Team
January 2022	Updated report submitted ^b	Evaluation Team

^a Delayed because unable to access NPD data due to Covid-19. Eventually agreed to proceed with publishing report without NPD data, the analyses of which to be later added. The initial version of the report excluding the FSM subgroup analysis was submitted in May 2021 and published in September 2021.

^b NPD data accessed and report updated to include results of analyses of NPD data.

Impact evaluation

Participant flow including losses and exclusions

Figure 2 shows the flow of participants and schools through the trial. In total, 437 schools were identified as meeting the recruitment criteria. They were consequently approached to take part in the study. Of these, 162 (37.1%) responded to the initial contact and were sent an information sheet and EOI form. Seventy-eight of the schools returned the EOI and were screened for eligibility; five were ineligible as they were already using, or had recently used, the Lexia programme. Consequently, 73 schools were sent a MOU; 59 returned it and so were willing and eligible to participate in the study, however, two schools withdrew prior to randomisation (one did not accept or sign the DSA and another had staffing issues). A total of 697 pupils were randomised within 57 schools: 348 to the intervention arm and 349 to the control arm. All 697 pupils completed the baseline testing and 620 pupils (89.0%) had complete primary outcome data (310 intervention, 310 control).

Table 8 shows the MDES at each stage of the study. At protocol, an MDES of 0.2 was calculated assuming 80% power, pre- and post-test correlation of 0.6, average cluster size of nine and attrition of 10%. At randomisation most of these

assumptions held but the observed average cluster size was higher than expected at 12.2 pupils per cluster. This increase allowed a slightly smaller MDES of 0.18. At analysis stage, we calculated the pre- and post-test correlation to be 0.75, higher than our original estimate of 0.6. The average cluster size in the analysis was 10.9. Although not included in our original sample size, an ICC at the school level was estimated to be 0.06. Including this further adjustment for clustering at school level allowed an MDES of 0.19 at analysis.

Figure 2: Participant flow diagram

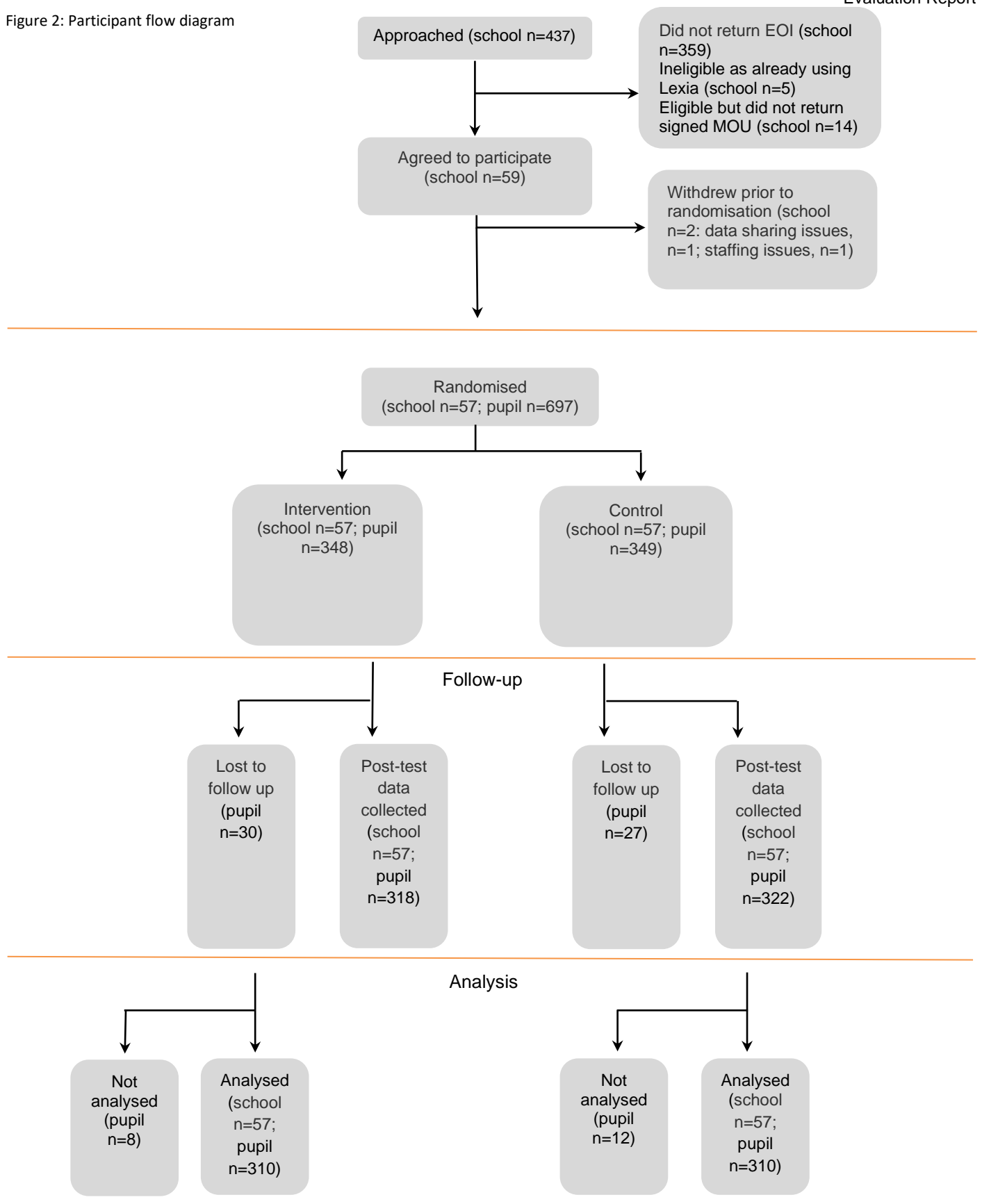


Table 8: Minimum detectable effect size at different stages

		Protocol		Randomisation		Analysis	
		Overall	FSM	Overall	FSM	Overall	FSM
MDES		0.20	0.36	0.18	0.31	0.19	0.29
Pre-test/post-test correlations	Level 1 (pupil)	0.6	0.6	0.6	0.6	0.75	0.74
	Level 2 (class)	-	-	-	-	-	-
	Level 3 (school)	-	-	-	-	-	-
Intracluster correlations (ICCs)	Level 2 (class)	N/A	N/A	N/A	N/A	N/A	N/A
	Level 3 (school)	N/A	N/A	N/A	N/A	0.06	0.075
Alpha		0.05	0.05	0.05	0.05	0.05	0.05
Power		0.8	0.8	0.8	0.8	0.8	0.8
One-sided or two-sided?		Two	Two	Two	Two	Two	Two
Average cluster size		9	3	12.2	4	10.9	4.2
Number of schools	Intervention	57	57	57	57	57	56
	Control	57	57	57	57	57	56
	Total:	57	57	57	57	57	56
Number of pupils	Intervention	257	86	348	114	310	114
	Control	257	86	349	114	310	119
	Total:	514	172	697	228	620	233

Attrition

The measure of prior attainment at baseline for the primary analysis was WRMT-R/NU composite reading score. This was available for all 697 pupils (100%). Pupils were included in the primary analysis model if they had data for the outcome (WRMT-III composite reading score) and covariates (WRMT-R/NU composite reading score, school). No data was missing for the covariates, however, 77 (11.0%) pupils had missing primary outcome data (Table 9). This led to the inclusion of 620 pupils (89.0% of those randomised) in the primary analysis. Across treatment groups the ratios (analysed to randomised) were 310:348 (89%) in the intervention arm and 310:349 (89%) for the control group.

The reasons for incomplete primary outcome data are presented in Table 10: 20 of 697 (2.9%) of pupils had an incomplete assessment and so did not have a WRMT-III composite reading score, 43 (6.2%) pupils left the school, six (0.9%) were absent at testing, two (0.3%) withdrew, and for six (0.9%) the reason for not completing the assessment was not stated. There were no great differences in the reasons for incomplete primary outcome data across arms.

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

		Intervention	Control	Total
Number of pupils	Randomised	348	349	697
	Analysed	310	310	620
Pupil attrition (from randomisation to analysis)	Number	38	39	77
	Percentage	10.9	11.2	11.0

Table 10: Assessment status of pupils recruited to the trial

Assessed status, n (%)	Intervention (N = 348)	Control (N = 349)	Total (N = 697)
Completed assessment	310 (89.1)	310 (88.8)	620 (89.0)
Incomplete assessment (no primary outcome score)	8 (2.3)	12 (3.4)	20 (2.9)
Left school	21 (6.0)	22 (6.3)	43 (6.2)
Absent at testing	5 (1.4)	1 (0.3)	6 (0.9)
Other reason	2 (0.6)	4 (1.1)	6 (0.9)
Withdrawn	2 (0.6)	0 (0.0)	2 (0.3)

Pupil and school characteristics

Characteristics of the 57 schools are presented in Table 11. As randomisation took place within school at the pupil level, school characteristics are presented overall. Most schools were in urban cities or towns (33.3%) or major conurbations (42.1%). A large proportion of schools were community schools (45.6%). The majority of schools were rated as 'good' in their most recent Ofsted inspection (61.4%). Nine schools (15.8%) had no Ofsted rating at the time the MOU was signed as they had recently become academy converters. Schools had a mean number of 418 pupils on role. A mean of 18.9% were receiving FSM and 29.9% had ever been on FSM. The mean percentage of SEN pupils with a statement or EHC plan across all schools was 1.1% and the mean percentage of SEN pupils with SEN support was 13.7%.

Our sample is fairly representative of the national level-means shown in the table. Our sample has more foundation schools (10.5%) compared to the national average (3%) and fewer voluntary aided (3.5% vs 17%) and voluntary controlled schools (5.3% vs 11%). The percentage of FSM pupils (18.9% compared to 15.7% national average) and percentage of SEN and those receiving SEN support (13.7% compared with 12.6% national average) are comparable to the national average.

Table 11: Baseline characteristics of schools recruited to the trial

School level (categorical)	National level mean	Overall	
		N (missing)	Count (%)
School setting*			
Rural town and fringe	N/A	57 (0)	6 (10.5)
Urban city and town	N/A	57 (0)	19 (33.3)
Urban major conurbation	N/A	57 (0)	24 (42.1)
Urban minor conurbation	N/A	57 (0)	8 (14.0)
Type of school			
Academy Converter	31%	57 (0)	14 (24.6)
Academy Sponsor Led		57 (0)	6 (10.5)
Community school	38%	57 (0)	26 (45.6)
Foundation school	3%	57 (0)	6 (10.5)
Voluntary aided school	17%	57 (0)	2 (3.5)
Voluntary controlled school	11%	57 (0)	3 (5.3)
Ofsted rating at time of MOU signed**			
Outstanding	18%	48 (9)	9 (18.8)
Good	69%	48 (9)	35 (72.9)
Requires Improvement	11%	48 (9)	4 (8.3)
Inadequate	3%	48(9)	0(0)
School level (continuous)***		N (missing)	Mean (SD)
Number on role	282	57 (0)	418.1 (139.3)

Percentage on FSM	15.7	57 (0)	18.9 (10.5)
Percentage ever on FSM	N/A	57 (0)	29.9 (14.2)
Percentage of SEN pupils with a statement or EHC plan	1.6	57 (0)	1.1 (1.2)
Percentage of SEN pupils with SEN support	12.6	57 (0)	13.7 (6.0)

* Source for data on school setting at national level cannot be located.

** Nine schools had no Ofsted rating at the time of signing the MOU.

*** National data sources: type of school, number on role, percentage FSM: DfE, 2019a; Ofsted rating: Ofsted, 2020; percentage SEN: DfE, 2021.

The baseline characteristics of the 697 pupils recruited to the trial and the 620 analysed in the trial are presented in Tables 12 and 13, respectively. Gender and FSM status were not available for 6 randomised pupils (intervention $n = 4$; control $n = 2$) because matching to NPD data could not be carried out. At randomisation, similar pupil characteristics were observed across trial arms. In the intervention group, 214/344 (62.2%) were male compared to 198/347 (57.1%) in the control arm, and 128/344 (37.2%) of pupils in the intervention arm had been eligible for FSM in the previous 6 years compared to 137/347 (39.5%) in the control arm. Our sample shows a higher percentage of boys (overall 59%) than the national average (51%), and also a higher percentage of FSM pupils than the national average. This may be due to the national average figures using a stricter definition of FSM (currently eligible for FSM) than the EVERFSM_6_P used (eligible for FSM in previous 6 years) which was taken from the 2019 Spring Census data. An additional explanation may be that we had a specific focus on schools with above average 'everFSM' pupils.

The mean pupil age in years was 6.36 in the intervention arm and 6.39 in the control arm. Pre-test scores were also very similar across arms, which is reflected in the low effect sizes for the difference. Pre-test Word Identification Standard Scores were collected for 343 of 348 (98.6%) intervention pupils and 342 of 349 (98.0%) control pupils. Pupils in the intervention arm scored an average of 110.83 (SD 11.40) and pupils in the control arm scored an average of 110.71 (SD 11.69). A Hedges' g effect size of 0.01 for the difference in mean scores between the groups was observed. Pre-test Word Attack Standard Scores were collected for 345 of 348 (99.1%) intervention pupils and 345 of 349 (98.9%) control pupils. Pupils in the intervention arm scored an average of 112.29 (SD 9.87) and pupils in the control arm scored an average of 112.76 (SD 9.66). A Hedges' g effect size of -0.05 was observed. Pre-test Passage Comprehension Standard Scores were collected for 344 of 348 (98.9%) intervention pupils and 347 of 349 (99.4%) control pupils. Pupils in the intervention arm scored an average of 98.25 (SD 10.88) and pupils in the control arm scored an average of 97.92 (SD 10.72). A Hedges' g effect size of 0.03 was observed. Pre-test Average Standard Scores were collected for all 348 (100%) intervention pupils and all 349 (100%) control pupils. Pupils in the intervention arm scored an average of 107.15 (SD 9.34) and pupils in the control arm scored an average of 107.06 (SD 9.47). A Hedges' g effect size of 0.01 was observed. Histograms of pre-test scores can be found in Appendix 12. Scores were normally distributed indicating no floor or ceiling effects.

Very similar patterns were observed in the pupils who were included in the primary analysis (Table 13).

Table 12: Baseline characteristics of pupils randomised to the trial

Pupil level (categorical)	National-level mean*	Intervention group ($n = 348$)		Control group ($n = 349$)	
		n/N (missing)	Count (%)	n/N (missing)	Count (%)

Gender						
Male	51.0%	344 (4)	214 (62.2)	347 (2)	198 (57.1)	
Female	49.0%	344 (4)	130 (37.8)	347 (2)	149 (42.9)	
FSM status**						
Not eligible for FSM	84.3%***	344 (4)	216 (62.8)	347 (2)	210 (60.5)	
Eligible for FSM	15.7%***	344 (4)	128 (37.2)	347 (2)	137 (39.5)	
Pupil level (continuous)		n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)	Effect size of difference
Age (years) at baseline assessment	-	348 (0)	6.36 (0.30)	349 (0)	6.39 (0.30)	-0.12
Word Identification Standard Score	-	343 (5)	110.83 (11.40)	342 (7)	110.71 (11.69)	0.01
Word Attack Standard Score	-	345 (3)	112.29 (9.87)	345 (4)	112.76 (9.66)	-0.05
Passage Comprehension Standard Score	-	344 (4)	98.25 (10.88)	347 (2)	97.92 (10.72)	0.03
Average Standard Score	-	348 (0)	107.15 (9.34)	349 (0)	107.06 (9.47)	0.01

* Data taken from www.gov.uk [*Gender, Percentage FSM*] (Schools, Pupils and Their Characteristics: January 2019 - GOV.UK, n.d.)

** 'Ever FSM' status is defined throughout as whether a pupil has been eligible for free school meals in the previous 6 years.

***National average of FSM group is defined as currently eligible for free school meals.

Table 13: Baseline characteristics of pupils analysed in the trial

Pupil level (categorical)	National- level mean*	Intervention group (n=310)		Control group (n=310)		
		n/N (missing)	Count (%)	n/N (missing)	Count (%)	
Gender						
Male	51.0%	310 (0)	195 (62.9)	310 (0)	179 (57.7)	
Female	49.0%	310 (0)	115 (37.1)	310 (0)	131 (42.3)	
FSM status**						
Never FSM	84.3%***	310 (0)	196 (63.2)	310 (0)	191 (61.6)	-

Ever FSM	15.7%***	310 (0)	114 (36.8)	310 (0)	119 (38.4)	
Pupil level (continuous)		n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)	Effect size of difference
Age (years) at baseline assessment		310 (0)	6.36 (0.30)	310 (0)	6.39 (0.29)	-0.09
Word Identification Standard Score		306 (4)	111.64 (10.77)	306 (4)	111.78 (10.74)	-0.01
Word Attack Standard Score		307 (3)	112.93 (9.32)	307 (3)	113.42 (8.96)	-0.05
Passage Comprehension Standard Score		306 (4)	98.90 (10.76)	308 (2)	98.82 (10.30)	0.01
Average Standard Score		310 (0)	107.85 (8.96)	310 (0)	107.97 (8.72)	-0.01

* Data taken from www.gov.uk [*Gender, Percentage FSM*] (Schools, Pupils and Their Characteristics: January 2019 - GOV.UK, n.d.)

** 'Ever FSM' status is defined throughout as whether a pupil has been eligible for free school meals in the previous 6 years.

***National average of FSM group is defined as currently eligible for free school meals.

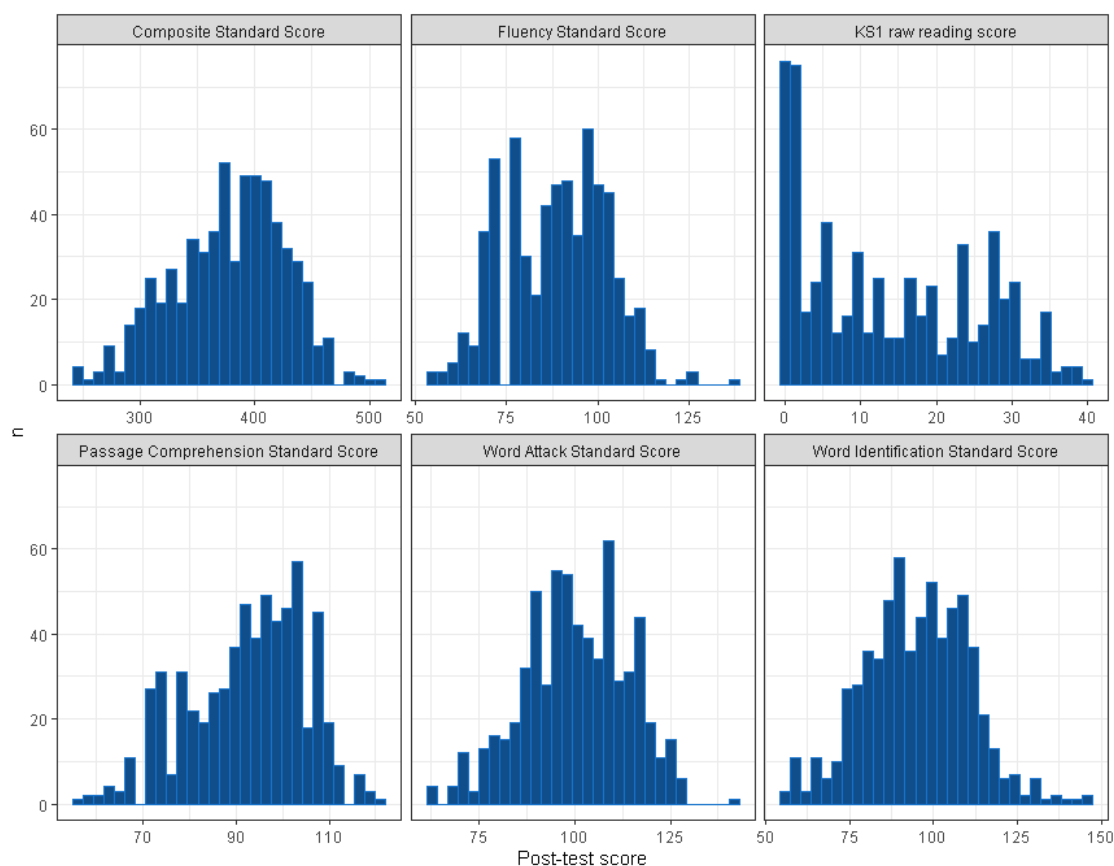
Outcomes and analysis

A descriptive summary of the post-test scores is presented in Table 14. The means were slightly higher in the intervention group for each of the outcomes. Histograms of the outcomes show they are all roughly normally distributed except for KS1 raw scores, which is positively skewed due to excess zeros and ones (Figure 3). As a further check of our model assumptions, the normality of residuals was checked using residual plots and residuals were shown to be normal (see Appendix 12).

Table 14: Summary of post-test scores for all pupils

Outcome	Intervention group		Control group	
	N (missing)	Mean (SD)	N (missing)	Mean (SD)
WRMT-III composite score	310 (38)	379.5 (49.2)	310 (39)	376.6 (48.2)
WRMT-III word identification score	318 (30)	95.8 (15.5)	322 (27)	94.5 (15.6)
WRMT-III word attack score	318 (30)	101.1 (13.5)	320 (29)	99.8 (13.8)
WRMT-III passage comprehension score	313 (35)	93.0 (12.2)	320 (29)	91.9 (12.5)
WRMT-III oral reading fluency score	314 (34)	89.0 (14.1)	313 (36)	88.6 (13.8)
KS1 raw reading score	300 (48)	13.8 (11.6)	308 (41)	13.8 (11.3)

Figure 3: Histograms of post-test scores



Primary analysis

WRMT-III composite reading scores were available for a total of 620 (89.0%) of the 697 randomised pupils (intervention $n = 310$, 89.1%; control $n = 310$, 88.8%, Table 15). A mean score of 379.5 (95% CI 374.0 to 385.0) was observed in the intervention arm and of 376.6 (95% CI 371.3 to 382.0) in the control arm. The adjusted mean difference is 3.63 (95% CI -1.34 to 8.57, Table 22). The estimated Hedges' g effect size is 0.11 in favour of the intervention (95% CI -0.04 to 0.27), which relates to approximately two months' additional progress, but this difference was not statistically significant (the 95% CIs contain 0, $p = 0.15$) and therefore could have occurred by chance. The total variance used to calculate the effect size was 1050.5, the sum of 66.4 (random variation between pupils, within-cluster variance) and 984.1 (heterogeneity between schools, between-cluster variance).

The assumption of normality of the standardised residuals was assessed using a QQ plot (Appendix 13: QQ plot for primary analysis model).

Secondary analysis

Secondary analysis results have been presented in Table 16.

Effect on word identification score

In total, word identification scores were available for 640 (91.8%) pupils (intervention $n = 318$, 91.4%; control $n = 322$, 91.7%). A mean of 95.8 (95% CI 94.1 to 97.5) was observed in the intervention arm and 94.5 (95% CI 92.9 to 96.2) in the control arm. The adjusted mean difference was 1.00 (95% CI -0.57 to 2.58, Table 22). The estimated Hedges' g effect size was 0.10 (95% CI -0.06 to 0.25), which relates to approximately two months' additional progress in the intervention group. No statistically significant difference was observed ($p = 0.21$). The total variance used to calculate the effect size was 107.9, the sum of 6.4 (random variation between pupils, within-cluster variance) and 101.6 (heterogeneity between schools, between-cluster variance).

Effect on word attack score

In total, word attack scores were available for 638 (91.5%) pupils (intervention n = 318, 91.4%; control n = 320, 91.7%). A mean of 101.1 (95% CI 99.6 to 102.6) was observed in the intervention arm and 99.8 (95% CI 98.3 to 101.3) in the control arm. The adjusted mean difference was 1.34 (95% CI -0.40 to 3.07) (Table 22). The estimated Hedges' g effect size was 0.12 (95% CI -0.04 to 0.27), which relates to approximately two months' additional progress in the intervention group. No statistically significant difference was observed ($p = 0.13$). The total variance used to calculate the effect size was 128.3, the sum of 5.0 (random variation between pupils, within-cluster variance) and 123.3 (heterogeneity between schools, between-cluster variance).

Effect on passage comprehension score

In total, passage comprehension scores were available for 633 (90.8%) pupils (intervention n = 313, 89.9%; control n = 320, 91.7%). A mean of 93.0 (95% CI 91.7 to 94.7) was observed in the intervention arm and 91.9 (95% CI 90.5 to 93.3) in the control arm. The adjusted mean difference was 0.92 (95% CI -0.68 to 2.51, (Table 22). The estimated Hedges' g effect size was 0.09 (95% CI -0.06 to 0.24), which relates to approximately one month's additional progress in the intervention group. No statistically significant difference was observed ($p = 0.26$). The total variance used to calculate the effect size was 111.7, the sum of 8.8 (random variation between pupils, within-cluster variance) and 102.9 (heterogeneity between schools, between-cluster variance).

Effect on reading fluency score

In total, reading fluency scores were available for 627 (90.0%) pupils (intervention n = 314, 90.2%; control n = 313, 89.7%). A mean of 89.0 (95% CI 87.4 to 90.6) was observed in the intervention arm and 88.6 (95% CI 87.1 to 90.1) in the control arm. The adjusted mean difference was 0.66 (95% CI -0.85 to 2.16, Table 22). The estimated Hedges' g effect size was 0.07 (95% CI -0.09 to 0.22), which relates to approximately one month's additional progress in the intervention group. No statistically significant difference was observed ($p = 0.39$). The total variance used to calculate the effect size was 95.9, the sum of 4.06 (random variation between pupils, within-cluster variance) and 91.8 (heterogeneity between schools, between-cluster variance).

Effect on KS1 raw reading score

In total, KS1 raw reading score was available for 608 (87.2%) pupils (intervention n = 300, 86.2%; control n = 308, 88.3%). A mean of 13.8 (95% CI 12.4 to 15.1) was observed in the intervention arm and 13.8 (95% CI 12.6 to 15.1) in the control arm. The adjusted mean difference was -0.06 (95% CI -1.77 to 1.10, Table 22). The estimated Hedges' g effect size was -0.04 (95% CI -0.19 to 0.12), so zero additional months' progress. No statistically significant difference was observed ($p = 0.64$). The total variance used to calculate the effect size was 90.7, the sum of 9.9 (random variation between pupils, within-cluster variance) and 80.8 (heterogeneity between schools, between-cluster variance).

Table 15: Primary 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
WRMT-III composite score	310 (38)	379.5 (374.0, 385.0)	310 (39)	376.6 (371.3, 382.0)	620 (310; 310)	0.11 (-0.04, 0.27)	0.15

Table 16: Secondary analyses

	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
WRMT-III word identification score	318 (30)	95.8 (94.1, 97.5)	322 (27)	94.5 (92.8, 96.2)	640 (318; 322)	0.096 (-0.055, 0.248)	0.213
WRMT-III word attack score	318 (30)	101.1 (99.6, 102.6)	320 (29)	99.8 (98.3, 101.3)	638 (318; 320)	0.118 (-0.035, 0.271)	0.131
WRMT-III passage comprehension score	313 (35)	93.0 (91.7, 94.4)	320 (29)	91.9 (90.5, 93.3)	633 (313; 320)	0.087 (-0.064, 0.237)	0.260
WRMT-III oral reading fluency score	314 (34)	89.0 (87.4, 90.6)	313 (36)	88.6 (87.1, 90.1)	627 (314; 313)	0.067 (-0.086, 0.221)	0.391
KS1 raw reading score	300 (48)	13.8 (12.4, 15.1)	308 (41)	13.8 (12.6, 15.1)	608 (300; 308)	-0.036 (-0.186, 0.115)	0.643

Analysis in the presence of non-compliance

Of the 348 pupils randomised to the intervention, 338 (97.1%) had data available on their use of the intervention; ten pupils from one school did not have such data as this school did not consent to this data being used for the evaluation. A summary of the use of the intervention is provided in Table 17. A median of 31 and inter quartile range (IQR) of 27 to 33 weeks with Lexia sessions was observed among the intervention pupils. The median number of total hours spent using Lexia was 34 (IQR 27 to 42); 269 of 338 pupils (79.6%) completed at least 60 minutes a week for 12 (non-consecutive) weeks (compliance definition one), and 314 of 338 (92.9%) completed a minimum of 720 minutes using Lexia (compliance definition two).

The results from the CACE analysis in Table 18 show similar effect sizes to the primary, intention-to-treat analysis. Using 'completed at least 60 minutes a week for a minimum of 12 (not necessarily consecutive) weeks' as the definition of a complier, an adjusted mean difference of 3.27 (95% CI -2.95 to 9.49) and a Hedges' g effect size of 0.10 (95% CI -0.09 to 0.30) were observed with no statistically significant difference ($p = 0.30$). Using 'completed at least 720 minutes in total' as the definition of a complier, an adjusted mean difference of 2.80 (95% CI -2.52 to 8.11) and a Hedges' g effect size of 0.09 (95% CI -0.08 to 0.25) were observed with no statistically significant difference ($p = 0.30$). Since compliance, as defined here, was reasonably high, these estimates do not differ largely from the ITT estimate (Hedges' g: 0.11, 95% CI: -0.04 to 0.27), as might be expected. Using the number of hours of the intervention received as a continuous measure (compliance definition three), an adjusted mean difference of 0.08 (95% CI -0.07 to 0.22) and a Hedges' g effect size of 0.002 (95% CI -0.002 to 0.007) were observed, for every increase in one hour of the intervention, with no statistically significant difference ($p = 0.30$). There is therefore no evidence that increased dose of the intervention translated to a greater benefit relative to teaching as usual.

Table 17: Summary of compliance in the intervention arm

Total weeks with Lexia sessions	
N	338
Mean (SD)	29.2 (6.7)
Median (IQR)	31 (27, 33)
Min, Max	(2, 36)
Total time (hours) spent using Lexia	
N	338
Mean (SD)	34 (14)
Median (IQR)	34 (27, 42)
Min, Max	(1,77)
Number completing at least 60 minutes a week for 12 non-consecutive weeks	
Yes, n (%)	269 (79.6)
No, n (%)	69 (20.4)
Number completing a minimum of 720 minutes in total	
Yes, n (%)	314 (92.9)
No, n (%)	24 (7.1)

Table 18: CACE analysis on the primary outcome

CACE analyses	Correlation ^a	F-test statistic	Adjusted mean difference (95% CI)	Residual standard error	Hedges' g (95% CI)	p-value
Completed at least 60 minutes a week for a minimum of 12 (non-consecutive) weeks	0.80	632	3.27 (-2.95, 9.49)	32.05	0.102 (-0.092, 0.296)	0.304
Completed at least 720 minutes in total	0.92	1951	2.80 (-2.52, 8.11)	32.00	0.087 (-0.079, 0.253)	0.303
Number of hours of intervention received	0.86	978	0.08 (-0.07, 0.22)	32.03	0.002 (-0.002, 0.007)	0.303

^a Correlation between the instrumental variable (randomly allocated group) and endogenous variable.

Missing data analysis

A mixed effects logistic regression model suggested that a 'missing at random' (MAR) assumption was reasonable, meaning multiple imputation by chained equations was an appropriate analysis approach (Group allocation: OR 0.98, 95% CI 0.59 to 1.63; pre-test score: OR 0.93, 95% CI 0.91 to 0.95).

To investigate the impact of missing data, the primary analysis was repeated using multiple imputation by chained equations (Table 19). The adjusted mean difference in WRMT-III composite reading score following multiple imputation was 3.09 (95% CI: -1.95, 8.14) and the Hedges' g effect size was 0.10 (95% CI -0.06 to 0.25); this difference is not statistically significant at the 5% significance level ($p = 0.23$).

The missing data analysis was repeated using multilevel multiple imputation. Similar results were observed: the adjusted mean difference in WRMT-III composite reading score following multiple imputation was 3.61 (95% CI: -1.49 to 8.71) and Hedges' g effect size was 0.11 (95% CI -0.05 to 0.27); this difference is not statistically significant at the 5% significance level ($p = 0.17$).

Table 19: Multiple imputation on primary outcome

	Adjusted mean difference (95% CI)	Residual standard error	Hedges' g (95% CI)	p-value
Multiple imputation by chained equations	3.09 (-1.95, 8.14)	32.40	0.095 (-0.060, 0.251)	0.229
Multiple imputation by chained equations with multilevel model	3.61 (-1.49, 8.71)	32.45	0.111 (-0.046, 0.268)	0.165

Subgroup analysis

Pupils eligible for FSM

Among the randomized pupils, 265 (38.0%) had ever been eligible for FSM (intervention $n = 128$; control $n = 137$), of which 233 (87.9%) had a complete WRMT-III composite reading score (intervention $n = 114$, 89.0%; control $n = 119$, 86.9%). The mean WRMT-III composite reading score for FSM eligible pupils was 373.7 (95% CI 363.2 to 384.1) in the intervention arm and 366.8 (95% CI 357.6 to 376.1) in the control arm (Table 20).

The mean WRMT-III composite reading score for never FSM eligible pupils was 382.9 (95% CI 376.6 to 389.1) in the intervention arm and 382.8 (95% CI 376.3 to 389.3) in the control arm; therefore, in general, FSM pupils tended to perform slightly worse than never FSM pupils.

Interaction model

When an interaction between allocation and FSM status was included in the primary model, the interaction was statistically significant at the 10% significance level (Hedges' g 0.31, 95% CI: -0.003, 0.63, $p = 0.053$).

Separate model for FSM pupils only

The adjusted mean difference in WRMT-III composite reading score between the intervention and control group among FSM eligible pupils is 9.47 (95% CI 0.44 to 18.52) (Table 21). The estimated Hedges' g effect size is 0.26 in favour of the intervention (95% CI 0.01 to 0.51), which relates to approximately 3 months' additional progress and is a statistically significant difference (p=0.04). The total variance used to calculate the effect size was 1310.6; the sum of 99.1 (random variation between pupils, within-cluster variance) and 1211.5 (heterogeneity between schools, between-cluster variance).

Table 20: Subgroup analysis among FSM pupils

	Unadjusted means				Effect size		
	Intervention group		Control group		Total n (intervention; control)	Hedges' g (95% CI)	p-value
Outcome	N (missing)	Mean (95% CI)	N (missing)	Mean (95% CI)			
Ever FSM	114 (14)	373.7 (363.2 , 384.1)	119 (18)	366.8 (357.6, 376.1)	233 (114; 119)	0.261 (0.012 , 0.511)	0.041

Table 21: Effect size estimation among FSM pupils

Outcome	Unadjusted differences in means	Adjusted differences in means (95 % CI)	Intervention group		Control group		Pooled variance
			N (missing)	Variance of outcome	N (missing)	Variance of outcome	
Ever FSM	6.86	9.47 (0.44,18.52)	114 (14)	NA	119 (18)	NA	1310.6

Estimation of effect sizes

The estimates used to calculate the effect sizes for each outcome are presented in Table 22.

Table 22: Estimate of effect sizes

Outcome	Unadjusted differences in means	Adjusted differences in means (95 % CI)	Intervention group		Control group		Pooled variance
			N (missing)	Variance of outcome	N (missing)	Variance of outcome	
WRMT-III composite score	2.84	3.63 (-1.34, 8.57)	310 (38)	NA	310 (39)	NA	1050.5

WRMT-III word identification score	1.28	1.00 (-0.57, 2.58)	318 (30)	NA	322 (27)	NA	107.9
WRMT-III word attack score	1.27	1.34 (-0.40, 3.07)	318 (30)	NA	320 (29)	NA	128.3
WRMT-III passage comprehension score	1.12	0.92 (-0.68, 2.51)	313 (35)	NA	320 (29)	NA	111.7
WRMT-III oral reading fluency score	0.40	0.66 (-0.85, 2.16)	314 (34)	NA	313 (36)	NA	95.9
KS1 raw reading Score	-0.06	-0.34 (-1.77, 1.10)	300 (48)	NA	308 (41)	NA	9.5

Estimation of ICC

The ICCs and 95% CIs associated with school (the cluster) are presented in Table 23 for the primary and secondary outcomes. The ICCs are within a similar range with the lowest ICC of 0.039 (WRMT-III word attack score) and highest of 0.109 (KS1 raw reading score).

Table 23: ICCs for primary and secondary outcomes

	Number of schools (clusters); n = 57	ICC (95% CI)
WRMT-III composite reading score	57	0.063 (0.010, 0.121)
WRMT-III word identification score	57	0.059 (0.011, 0.117)
WRMT-III word attack score	57	0.039 (0.000, 0.092)
WRMT-III passage comprehension score	57	0.078 (0.024, 0.142)
WRMT-III oral reading fluency score	57	0.042 (0.000, 0.098)
KS1 raw reading score	55	0.109 (0.046, 0.182)

Implementation and process evaluation results

As discussed above, the IPE research questions (Table 6) were:

- 1 What is the relationship between the fidelity of the intervention and the impact on pupil outcomes?
- 2 How much variability occurred in implementation of the intervention across different settings in respect of:
 - a) dosage (number and frequency of overall sessions)?
 - b) school factors, such as physical space/place of intervention?
 - c) teacher factors, including supervision of the sessions and use of paper-based resources?
 - d) potential or actual perceived barriers to implementation?
3. The reach of the intervention (including use of training and support provided)?
4. The nature of teaching as usual (that is, the control conditions)? What support was offered to those pupils not allocated to the intervention group?

This section discusses:

- training and support—IPE RQs 1 and 3: reach and fidelity;

- compliance—IPE RQ 2: dosage;
- implementation and fidelity—IPE RQs 1 and 2: implementation and fidelity;
- reach of the programme—IPE RQ 3: reach;
- usual (non-Lexia) reading practices—IPE RQ 4; control conditions; and
- additional considerations—IPE RQs 1, 2, and 3: perceived relationship with fidelity and pupil outcomes, variability and reach.

The main findings from the IPE are:

- The training was completed by SMTs, Year 2 teachers, and TAs, fitting in with the whole-school approach envisaged in the logic model.
- There were observed effects in the impact evaluation relating to dosage and outcomes and compliance was high in this study.
- Delivery was mainly carried out by TAs and, despite SMT involvement in training, responsibility for delivery and monitoring of the programme fell to the TAs with support from Year 2 teachers.
- Some schools struggled with managing delivery in terms of providing extra space for programme delivery (which is an issue within primary schools). Over half of all schools that responded to the post-intervention survey reported no issues dealing with IT provision and connectivity.
- Reports were generally accessed at least weekly by schools (69%) but mainly by the programme deliverer (the TA: 39%) or the Year 2 teacher (32%) rather than being incorporated within school practices more widely (that is, accessed by members of the SMT).
- There were issues with the reach of the programme with some concerns surrounding selection of the evaluation pupils and the suitability of the programme for some pupils, particularly for SEND pupils.
- Pupils generally enjoyed the programme and felt it gave them confidence in their learning.
- Lexia provision tended to be fitted into the school day in a way which meant that intervention pupils in most schools also received the normal provision for struggling readers on top of the Lexia programme.

Some sections cross over research questions but provide an overall research narrative, particularly in terms of some schools' considerations and experiences. Consequently, this was perceived to be the best way in which to organise the reporting with the results drawn together in the overall discussion.

Training and support

As indicated in the logic model and TIDieR table, three webinar training sessions were designed to take place for each school involving Year 2 teachers, TAs, and a senior member of staff. Unfortunately, attendance at training was not collected at the point of delivery. However, from across 50 schools, 70 (70%) of the 98 post-test survey respondents reported undertaking at least one of the three webinar training sessions (Table 24). Thirteen more (13%) reported receiving the training by other means, for example, 'telephone contact with LexiaUK representative' (LexiaUK tended to use a screen sharing solution (visual) in conjunction with the phone (audio)). Thirteen (13%) of the remaining 15 respondents had neither delivered the programme nor received any training and the remaining two (2%) said they had not received any training but had nevertheless delivered the programme. In addition, LexiaUK indicated that at least three members of staff attended each session, including a senior member of staff from each school, according to email correspondence between the evaluation team and LexiaUK.

Thirty eight (54%) of the 70 respondents who attended at least one of the webinar training sessions actually facilitated the delivery of the Lexia programme to the EEF study pupils. An additional 15 respondents also delivered the programme, 8 (11%) of whom reported having been trained by a colleague or senior member of the teaching staff who did attend the sessions. Nearly three-quarters of the Year 2 teacher respondents (42, 74%) undertook some or all of the training, but only around a quarter of those (11, 26%) reported actually delivering the programme, whereas nearly two-thirds of the TA respondents undertook some or all of the training (22, 61%) all of whom delivered the programme. Whilst it is not possible to determine exact numbers, there was evidently a tendency in many schools for the teachers to oversee the delivery and to be responsible for the more administrative or decision-making aspects of the programme, such as the paper reports and any difficulties that individual pupils were experiencing, whilst the TA ran the classes.

This fits with the logic model of the programme enacting change within the school and providing an integrated approach with teachers aware of the programme in order to best support the children selected to benefit.

Of the 53 respondents who actually delivered the programme, 35 (66%) had attended at least two of the webinar training sessions, three (6%) had attended one session, and 15 (28%) had not undertaken any of the webinar training.

Table 24: Attendance at training (n = 98)

Survey respondents	Delivered the Lexia programme or not n (%)	Attended all 3 webinar sessions n (%)	Attended 2 webinar sessions n (%)	Attended 1 webinar session n (%)	Did not attend the webinar training n (%)
Yr 2 teachers (total = 57)	Yes 16 (28)	4 (7)	7 (12)	0 (0)	5 (9)
	No 41 (72)	12 (21)	11 (19)	8 (14)	10 (18)
Teaching assistants (TAs) (total = 36)	Yes 32 (89)	11 (31)	8 (22)	3 (8)	10 (28)
	No 4 (11)	0 (0)	1 (3)	0 (0)	3 (8)
Others (for example, heads/learning leaders; total = 5)	Yes 5 (100)	3 (60)	2 (40)	0 (0)	0 (0)
	No 0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
TOTALS	Yes 53	18	17	3	15
	No 45	12	12	8	13

Of the 53 respondents who delivered the programme, all but four (92%) said that the training they received gave them sufficient information to enable them to deliver it. Two of the four who disagreed had undertaken the webinar training the other two (one teacher and one TA) said they had not been trained.

Survey respondents who undertook the webinar training were asked to rate the training they received (those who were trained by other means did not provide a rating). Nobody was actively dissatisfied with the training and all but five respondents (7%) were 'satisfied' or 'very satisfied'. Table 25 shows the satisfaction rating among those who attended the webinar training who also said they delivered the programme (38 respondents).

Table 25: Training satisfaction rating among programme deliverers (38 individual respondents)

	Very satisfied n (%)	Satisfied n (%)	Neutral n (%)	Dissatisfied n (%)	Very dissatisfied n (%)
Programme deliverers who attended the training	18 (26)	17 (24)	3 (4)	0 (0%)	0 (0%)

Two-thirds (18) of those who were 'very satisfied' with the training highlighted the clarity or thoroughness of the material or the clear and supportive communication of the trainer. The remaining nine highlighted related benefits, such as gaining the confidence necessary to deliver the programme, and that it was fun, and friendly.

For those who gave the training a 'satisfied' rating, over two-thirds of the comments (n = 26) were also positive ('good', 'adequate', or 'sufficient for purpose'). Others mentioned minor shortcomings or criticisms, such as the training feeling fast-paced or rushed in parts (n = 3), needing more guidance or information (n = 2), and the delivery mechanism (in this case the webinar) having limitations (n = 1). The remainder, together with those who were neither satisfied nor dissatisfied, left no comment.

Only 11 survey respondents took the opportunity to offer any suggestions for improvement of the training and their comments were wide ranging. These encompassed a desire for:

- more support—for example, online, for reading data ('perhaps each week'), and after three to four weeks (n = 4);
- more information about pupil eligibility (n = 2);
- more written material to support the webinars—for example, emailing keynote presentations and instructions so that additional staff could be trained (n = 2);
- more face-to-face input—for example, someone coming into schools sometimes as technology does not always work (n = 2); and

- adaptation, where necessary, of terminology to suit U.K. practice (n = 1).

The interview respondents were also very positive on the whole regarding the training received. For example:

'[The training] was very good. It was very informative and we could do it at our own pace, have a conversation if we had to clarify anything' (Year 2 teacher).

'[The training] was really good actually ... I was concerned with the mechanics of it, sitting at a computer, but it was good and very clear' (TA).

Where problems were identified (by four respondents), they all related to the amount there was to learn and the lack of time to embed the content. One teacher wished she could have had more time to become more familiar with the games, another said it was 'much easier to pick up when you're able to practise it', and a teacher and the TA for her class both thought there was a lot to take in, and remember:

'I think I would have preferred more time to go away and have a look at the games' (Year 2 teacher).

'I don't think you take as much in when you're watching. I think it would have been better to have it in front of us and to be able to explore it while listening and watching. I found it very difficult to follow and then later go on the programme and still remember everything' (Year 2 Teacher).

Compliance

Compliance was defined as completing at least 60 minutes a week for a minimum of 12 non-consecutive weeks or completing a minimum of 720 minutes over a 24-week period. Tables 17 and 18 above provide detail on the CACE analysis using online usage data. As seen, approximately 80% of pupils completed at least 60 minutes a week for 12 non-consecutive weeks and approximately 93% of pupils completed a minimum of 720 minutes using Lexia. There were observed effects for compliance, although they were not statistically significant. It does, however, suggest that compliance did result in higher, albeit non-significant, outcomes for pupils, which means these may have happened by chance.

In terms of the observations, the setup of the lessons only took one or two minutes—although the delivery team had specified ten minutes for setting-up and ending the lessons, hence the 30 specified minutes with 20 minutes of delivery—mainly because the programme deliverers had the room and equipment set up prior to the children arriving and all children appeared familiar with the routine. The computers (laptops, PCs, and tablets) were already switched on and most children then knew how to log in to the software. One school had cards for each child reminding them of their login details but otherwise children knew their own details with only six children across five of the observed schools needing teacher assistance in logging into the software. In no case was this felt by the researchers to be because the child was not familiar with the routine but was more due to the need for additional scaffolding (see section on 'Reach' below).

Further details on implementation and fidelity—other facets of compliance—are provided in the next section.

Implementation and fidelity

Given that the online element is designed to provide personalised learning for each student it was recommended that no further adaptations were made to the delivery although it was felt important that the IPE assessed the conditions and nature of implementation in schools. Consequently, a number of questions were asked in the post-intervention survey in order to explore any variability in the delivery of the online element to the Year 2 pupils participating in the intervention. 'Fidelity' was defined as the online student activities taking place outside of the classroom with compliance (as defined above), with full staff training, accessing the online reporting data at least once a week, and using headphones to cut-out external distractions for pupils (as indicated in Table 3). From some of the responses, however, it is clear that at least some respondents failed to distinguish between the part of the programme delivery that was being evaluated and either the embedding phase (summer 2018) or the additional delivery that was licensed by the agreement but that fell outside the evaluation—a misunderstanding, the extent of which cannot be accurately ascertained. For example, one respondent replied to the question of how many children were in the group by saying 'six on the study and 24 doing Lexia', but another answered the same question simply by saying '20'. The likelihood, therefore, is that the latter is quantifying all the pupils doing Lexia rather than just those within the evaluation. Where answers could not be fully verified, they were treated as missing unless specifically referred to in the reporting that follows. The allocated group

size for the intervention in this evaluation was determined to be six to seven Year 2 pupils which is different from normal, whole-school implementation conditions (see below). For the observation schools, with the exception of one school (as detailed below) the average number of pupils in a session was six (ranging from four to six per sessions).

Who delivered the programme

Fifty-three (54%) of the 98 post-intervention survey respondents said that they delivered the programme to the EEF study pupils (from 40 schools). Only 16 (28%) of the Year 2 teacher respondents said they delivered the programme whereas all the TA respondents (n = 34) said they delivered the programme, as did all seven of those in the other roles. TAs, therefore, made up nearly two thirds (64%) of those who delivered the programme although in the observations the delivery was entirely by TAs. We suspect, therefore, that Year 2 teachers indicated they delivered the programme in the majority of cases because they facilitated its delivery.

Respondents were also asked who primarily delivered the programme in their school. As seen in Table 26, in 44 of the 52 schools represented (85%), respondents reported that it was the TA who did so. In all ten of the observation schools a TA delivered the programme.

Table 26: Who primarily delivered the Lexia programme in your school?

Role	n (%)
Teaching assistant only	44 (85%)
Teaching assistant and class teacher	4 (8%)
Class teacher only	2 (4%)
Other	2 (4%)

n = 52 schools.

Given that TAs were the main deliverers of the programme the researchers asked how they came to be involved. For the most part it appeared to be a pragmatic decision based primarily on availability.

TA: 'Why was I chosen?'

Asst. head: 'Well because you were available, I hate to say it that way but ...'

TA: 'And I'm happy to deliver it. Because I'm, like, the third TA—they use me wherever they need me so that's why.'

'She is my class TA and, because we've got other interventions going around in school [and] me needing to be in certain places ... [she's] the one person that we could release all the time that would be consistent for them, as well ... she knows all the children and they'll ask for help if they need it' (Year 2 teacher)

'... because she's available (laughter). She doesn't get pulled out to do other things' (Year 2 teacher).

'Well [TA] is our general LSA (Learning Support Assistant) [whereas] some of the other one to ones have specific responsibilities during that time [and] she was a good match for what we needed' (assistant headteacher).

Equipment used

The most frequently listed device used for accessing Lexia was a tablet with 23 schools (44%) listing this as the only equipment used. Laptops were the only device used in 14 schools (27%) and nine schools (17%) listed PCs as the only equipment used. The remaining six schools used a combination of the three devices (Table 27).

In the observation schools, four used laptops (although one school supplemented this with one PC), three used tablets, and two used PCs only. Whilst overall the children in the observation schools usually used headphones, in one school

this was not usual practice but they did use them for the observation. In one other school, two of the children did not use headphones so were sat further apart from the other children to avoid distracting them.

Table 27: Equipment used for delivery of the Lexia software

Equipment	n (%)
Tablets only	23 (44)
Laptops only	14 (27)
PCs only	9 (17)
Combination of tablets, laptops, or PCs	6 (12)

n = 52.

Where the programme was delivered

Forty-one schools (85% of the 48 schools that answered this question clearly) delivered the programme outside the classroom and seven (13%) delivered it inside. Alternative locations to the classroom included the school computer suite (n = 14, 27%), a small study room (n = 13, 25%), and the library (n = 7, 13%) (Table 28).

Table 28: Where the online activities were delivered

Location	n (%)	
In the classroom (i.e., alongside the lessons being delivered for the rest of the class)	7 (13)	
Outside the classroom	Small study room	13 (25)
	Computer suite of PCs	12 (23)
	Computer suite of PCs (combined with library or small study room)	2 (4)
	Library	7 (13)
	Other (e.g., other empty classroom, shared area, corridor, lower hall)	7 (14)
Missing	4 (8)	

n = 52.

Three of the ten observed sessions in the case study schools took place in the library, two in the IT suite, two in a small learning/intervention room, and one in a separate classroom. However, one of the sessions took place in a large hall where three or four other groups were working and one in a shared communal space between classrooms. In addition, one of the library delivery sessions observed had been moved into that space especially for the research visit (despite being asked to continue as usual to avoid disruption) as sessions usually took place across the three Year 2 classrooms, during lessons, with the TA circulating between the three rooms to provide support. However, whereas three of the observations were rated as taking place in more 'noisy' environments than the other seven (rated as 'quiet') in all but one the children were wearing headphones and in no cases did the children appear to be distracted. In addition, space seemed to be adequate in all cases, even when children were placed quite close together (as this seemed more conducive to TA supervision).

When the programme was delivered

Of the 52 schools represented in the post-intervention survey, together with two interview schools that also provided relevant data, 17 (31%) reported delivering the sessions at different times during the week alongside both assembly or lesson time; 25 (46%) delivered the sessions during lesson time only, and ten (18%) said they delivered the sessions alongside assembly only. One school delivered the sessions before the school day started and experienced problems with one or two of the pupils struggling to arrive on time.

Many of the interview respondents referred to their school's thinking in relation to scheduling the delivery of the programme. Whilst schools made different choices between the beginning, middle, or end of the day, in all cases, the common objective was to minimise the Lexia pupils' absence from their normal curriculum.

'The Lexia session needs to finish by 1.30 ... we've tried to make sure that we're keeping them in class as much as we can really' (assistant head).

'They're only missing assembly ... we do try and timetable so that they are not missing out' (Year 2 teacher).

'We try and schedule it at the end of the day ... to make sure they miss as little as possible' (TA).

The TA from one school that had chosen to run the sessions at 2pm, did acknowledge:

'One of the problems we're having is that they're not doing the independent part of that lesson. It's taking them away from that.'

For the observation schools, six scheduled their lessons in the afternoon, three in the morning, and one varied according to the day of the week.

Accessing the programme

Of the 53 survey respondents (from 42 schools) who delivered the programme, just over one third (n=19, 36%) in 19 schools said they experienced some issues accessing the program (Table 29).

Table 29: Ease of access to the programme

Did you experience any issues accessing the Lexia programme?	n (%)
Yes	19 (36)
No	27 (51)
Missing	7 (13)

n = 53 individual respondents. The kinds of issues experienced were mostly technical problems of one sort or another, either related to the school equipment or (more commonly) to the software. These included:

- problems loading the programme or logging on (n = 8);
- children being logged out suddenly (n = 4);
- school equipment issues (n = 4);
- programme freezing issues (n = 2);
- sound issues (n = 2); and
- issues with the time-tracking function (which tracks the speed of children's responses within the programme) (n = 2).

Thirteen (79%) of the 19 survey respondents who experienced issues said that they accessed additional Lexia support, both by email and telephone, where it was felt necessary. From the interview data (available for eight schools), respondents in three of these had not felt the need to ask for additional support. The reasons respondents in the remaining five schools had asked for further support related to administrative (three schools), as well as programme (two schools) and technical (two schools) concerns. All the requests for administrative support related to the need to withdraw children who were participating in the study or make changes to other licensed children within the school. The programme and technical concerns were more varied.

'I phoned up and used the hotline because we wanted to change some of the children ... so they turned me over to one of the administrators, so I have kind of contacted them for help that way' (assistant head).

'[We've asked for help with] tracking progress, looking at how we can evidence progress of Lexia, which we're still kind of looking at because the children are at different starting points, not just for the study children, for the other children with the extra licences, too' (Year 2 teacher).

'I had a couple of hiccups with an error on it once and had to email the support group ... and they were pretty quick actually' (TA).

'The other [support query] was, and that happened this morning a couple of times, is the sound will drop off, you can hear it really faintly ... and you've got to log out, log back in for it to solve the problem' (TA).

Respondents were very positive about the quality and efficiency of the support available.

'Whatever we need, they're always there to support us' (Year 2 teacher).

'That's literally a two-hour turnaround, that's how quick, you know, if I asked them a question at four o'clock on an afternoon then the answer is there the next morning or if I want to add a student I've added them at 9.00am and they're there by 11.00am, so from that point of view that's been perfect' (TA).

Use of the Lexia reports

Schools were asked to access the Lexia reports at least once a week during the intervention period to check pupils' progress and utilise printable resources if appropriate. Respondents in 48 schools (92%) said they accessed the reports provided by the Lexia software. Respondents in just under two-thirds of schools (n = 33, 63%) said they accessed the reports weekly and 15% (n = 8) said monthly (Table 30). Forty-four of the 48 schools that said they accessed the reports also responded to the question of who accessed them. This was predominantly the programme deliverers (39%) which, as seen above, was predominantly TAs, followed closely by class teachers (32%) (Table 31).

Table 30: Frequency of access to Lexia reports

	Schools n (%)
Daily	3 (6)
Weekly	33 (63)
Fortnightly	2 (4)
Monthly	8 (15)
As needed	1 (2)
In July as an end of year report	1 (2)
Did not access/did not know	2/2 (8)

n = 52 schools.

Table 31: Who accessed the reports?

	Schools n (%)
Class teachers	14 (32)
Programme deliverer(s),	17 (39)
Senior staff, for example, head or assistant headteacher, Lexia lead	5 (11)
Those delivering the programme and senior staff	4 (9)
Those delivering the programme and specialist SEN staff	4 (9)
Missing	8 (15)

n = 52 schools.

Respondents were asked in both the survey and the interviews how the reports were used. In the survey this was a free text question and, superficially, respondents varied widely in their responses. However, aside from five schools that did not provide a response, at least one of the respondents in the remaining 47 schools essentially described the use of the reports to monitor the progress of the children and, in 12 of these, the respondent then went on to describe the monitoring as a mechanism and guide for the subsequent adaptation of their learning strategies. Other survey responses included the use of reports to print off certificates (four schools), to give feedback to the SENCO (one school), and checking on time covered (one school).

Responses from the face to face interviews ranged from not using the reports at all (one school) to logging on 'pretty much every session to check how they're doing' (TA). Typically, respondents in the interview schools echoed the primary use of the reports as a monitoring tool with a view, in some cases, to adapting learning strategies either for individual children or a group as a result.

'I look at the [reports] to see where they're struggling within the programme because, obviously, it's still linking into their phonics, although they still do have extra phonics interventions, anyway ... so then it's helping them again in another way' (Year 2 teacher)

I keep tabs on the times and also just out of interest I look at how many attempts they've done at various things just for my own interest because I watch them and just see whether my observations, are they real or not. (TA)

I do have a little look [to] see if they're spending too long on an activity, how close they are to finishing ... sometimes I think, "Oh, you've been on that for ages," when you need a certificate to boost yourself. (TA)

I check it every Friday ... basically to see how they're getting on, and it tells me whether they need extra support or if they need some written sheets ... (TA)

'So I look at them perhaps once every three weeks to a half term. [The head] looks at them weekly. He has quite a good overview, and then he'll share information with us so, for example, there's a percentage that lets you know the prediction and so one of the children is like on a 94% ... whereas somebody else is like at 1% ... so yes, he looks at it every week' (Year 2 teacher).

'I log on regularly during the week just to check that they've met their times because occasionally things will change in the timetable. But making sure, then just reminding teachers [to] plan in that time for the children to use Lexia ... I look at it in terms of also strategically printing the reports and the progress as well as a part of my literacy review in autumn term ... and also just informing the rest of the senior leaders about what kind of things the children are doing' (assistant head).

Survey respondents in two schools expressed some reservations about the reports. One teacher said that the TA 'didn't really understand how to use them to the best of their ability', and another said she used them as a general guide but that 'they were hard to understand'.

Use of the Lexia paper-based resources

As the assistant head indicated above, the report data was used to inform the use of the paper-based resources, which take the form of worksheets for the children to complete or scripted lesson plans to assist TAs and teachers in instruction where pupils struggled. Comments from survey respondents did not address this point, but other interview respondents also indicated that their use of the paper-based resources was guided in this way, so it is reasonable to assume that this was a common practice.

'I will try to have a look at the graphs on the reports ... and Lexia will say there is a problem with this [so I] print those sheets off and [the class teachers] sit with them, and sometimes it's a bit of a one to one' (TA).

'When it comes to printing out sheets, it tends to depend on what the [report] graphs are doing because if everything's going smoothly, we just leave it ... but then sometimes [I'll] notice [that] quite a few of them are struggling ... and I'll start printing out some sheets again' (TA).

'[The head] looks at [the reports] weekly ... and then if there's lessons that flag up using the printed resources ... he brings it to our attention' (Year 2 teacher).

This latter comment, however, was one of the few comments relating to SMT involvement in programme delivery.

As seen in Table 32, just over a quarter of schools (n = 14) represented in the survey used the paper-based resources on a regular basis (mostly weekly but also daily, fortnightly, and monthly in one school each) whereas another quarter used them as and when a child was perceived to need additional help.

Table 32: Frequency of use of paper-based Lexia resources

	Schools n (%)
Weekly	11 (21)
Other regular frequency—e.g. daily (n = 1), fortnightly (n = 1) or monthly (n = 1)	3 (6)
Whenever a child needed extra help (n = 13) or when flagged by the programme (n = 3)	16 (31)
Other ('occasionally' (n = 1), twice over the duration of the programme as a whole (n = 1))	2 (4)
Never	12 (23)
Conflicting responses from same school/missing	8 (15)

n = 52 schools.

Respondents in nearly a quarter of schools did not use the paper-based resources at all. Whilst survey data is not available to explain this in any more detail, respondents in three of the interview schools referred to issues of time and other practicalities, which may well have been the case in other schools.

Teacher: 'The paper [resources are] quite basic and a bit repetitive, but that is what they need ... maybe a bit more time in the paperwork activities.'

TA: 'It's nice having the paperwork, but sometimes it is tricky delivering ... if a child is struggling, they might need the paperwork but the others don't ... you would have to plan [a class] when they are all doing some form of paperwork ... but ideally you need to do that child's paperwork there and then [but there's] not always time, especially if an adult needs to sit with that one child' (Year 2 teacher and TA).

'The trouble is we're struggling to actually: it'll say they have to have a lesson on the paper sheets and ... we don't have time really to fit it in because they're doing this and missing so much, so I mark it as delivered but then, when I've seen that they're struggling, I tend to sit with them and we have headphones off and we talk about it and we do it together on the screen rather than on the paper. We just can't find the time to fit it all in' (TA).

'No, we don't use those ... because [of] the time it takes to get on, print them out' (Year 2 teacher).

Excluding the schools in which respondents said they did not use the paper-based Lexia resources (n = 12), two-thirds of the remaining 40 schools said that it was the teachers or the teaching assistants who accessed them (Table 33).

Table 33: Who accessed the paper-based Lexia resources?

	Schools n (%)
Teachers or teaching assistants	30 (58)
SEN teachers (two who did not deliver the programme, one who co-delivered)	3 (6)
Not applicable (12 schools) or missing (7 schools)	19 (37)

n = 52 schools.

Although the programme can be provided for use at home, for the sake of the evaluation, schools were asked to ensure that it was only delivered in school. This was partly to ensure that pupils without wider access to IT were not

disadvantaged and partly to facilitate greater consistency of implementation fidelity. As Table 34 shows, whilst most of the schools that used the paper-based resources provided them to children to work through in class, 11 schools (22% of the survey sample) provided them, either exclusively or as well, to children for homework. Whilst this practice did not compromise access to the programme as a whole exclusively within the school, it did have the potential to distort the consistency of implementation fidelity. There is no evidence that these resources were provided to control children and it seems unlikely as the paper-based resources were directly related to the stage of the programme each intervention child had reached.

Table 34: How were the paper-based resources used?

Provided to children to work through in class	Provided to children for homework	Provided to teachers	Schools n (%)
✓			20 (38)
	✓		5 (10)
✓	✓		5 (10)
✓		✓	4 (8)
✓	✓	✓	1 (2)
		✓	1 (2)
Not applicable or missing			16 (30)

n = 52 schools.

The usage of paper-based resources is illustrated by the interview responses below:

'I would use the paper resources one to one during the Lexia session while the other children were on the iPads. I would pass on any difficulties to teachers so that resources could be provided to the child by the class teacher. One particular child struggled to read the tricky high frequency words so these were taken from the programme and used as flashcards in class and sent home as part of his homework' (TA).

'I use [the paper resources] for when they're struggling ... I'm not a specialist in literacy [so] I use those' (TA).

'Well, it actually prints off a set of skills sheet every time there's a certificate and I send those home with the children and they do those at home with their parents. If it sends me a lesson then I always print that off and deliver it here' (TA).

'They are [also] quite keen to take the sheets home ... they love doing Lexia' (TA).

'All children in the group worked with them during the time we allocated to the Lexia programme. Especially children struggling to access the work' (higher level teaching assistant, HLTA).

'They were used during Lexia sessions when children were struggling with concepts and needed specific tuition' (dyslexia specialist teacher).

Programme reach

The reach of the programme training was reported as predominantly to TAs and Year 2 teachers, although the team is unsure about the extent of senior leadership participation and few headteachers were described as taking an active interest in the reports on usage.

In terms of pupils to target, the programme can be used with struggling readers or more widely. This evaluation chose to focus on struggling readers. However, as indicated in the Participant Selection section, the evaluation team, in conjunction with the EEF, altered the criteria for struggling readers to be within school rather than across schools. Some

issues were reported around the suitability of participants selected due to this change in criteria, and some withdrawals or adaptations were made to delivery due to school perceptions of the suitability of the programme for some pupils.

Changes in delivery for some pupils were reported in ten (19%) of the 52 schools represented in the survey. In seven schools, the changes involved pupils who withdrew from the sessions completely and in three, the pupils continued to take part but the delivery had to be modified in some way, for example, by focusing on particular aspects of the programme or by adapting the length or number of sessions. Five schools identified 'learning difficulties' as the main reason for either withdrawal (three schools) or adaptation (two); 'persistent absence' was given as the main reason in three schools (in one school, the sessions were run before the normal school day started and the parents failed to bring the child in early enough) and in the remaining two schools, 'behavioural difficulties' were identified as a reason for withdrawal in one and 'emotional difficulties' in the other.

Respondents in 17 schools (33%) thought there were certain pupils for whom the programme would not be suitable. Most commonly, the programme was considered unsuitable for pupils whose SEND (Special Educational Needs and Disability) needs were too severe or complex (n = 9 respondents). Respondents also felt that the programme was potentially unsuitable for children whose ability levels were too low (n = 6), those who needed one to one support (n = 4), those whose EAL levels were too low (n = 3), and one respondent thought it was unsuitable for those with behavioural issues.

Those who needed one to one support were clearly likely to be pupils with one or more of the needs identified in the other characteristics (for example, SEND, ability, EAL, or behavioural issues). It is important to note a particular difficulty that the programme presented in a small number of schools where the introduction of the Lexia programme constrained (or eliminated) the resources available for one to one support for other struggling readers, as the following comments illustrate:

'One child was identified by Lexia but didn't recognise any letters so found it hard to access' (Year 2 teacher).

'[It was] difficult to administer the programme and offer one to one support—one to one reading with a child has proved more beneficial than this learning platform' (Year 2 teacher).

'One of the SEND learners couldn't retain and soon reached a level in which they couldn't progress. We did not have the opportunity to utilise paper lessons due to the time that was required for Lexia already to be able to do. Each child had different targets that needed addressing and we did not have the staff to deliver' (Year 2 teacher).

Table 35 shows which groups of pupils survey respondents felt they would particularly recommend for the programme.

Table 35: Groups of pupils survey respondents would particularly recommend for this programme

	N (%)
Would recommend for whole school—all pupils	15 (15)
Struggling readers only	25 (26)
Struggling readers and EAL pupils	35 (36)
EAL pupils only	4 (4)
Non-struggling and non-EAL pupils	8 (8)
Other: SEND only: 1 Just below expected level, i.e. not SEND: 1 Year 1 only: 2 Year 2 only: 1 Year 3 upwards: 1	6 (6)
Missing	5 (5)

n = 98 individual respondents.

Whilst the responses to the questions of suitability do not give a completely clear picture, indications are that respondents regarded the programme as most suitable for struggling readers or EAL pupils performing—as one respondent put it—‘just below the expected level’.

During the observations it was noted that whilst all TAs circulated the room and provided help—both requested and unrequested—in some cases (three schools) the focus tended to be on particular children who appeared to need additional support which could be up to 11 times in one 20- to 25-minute lesson, or the TA positioning themselves next to a particular child during most of the session and only circulating briefly. This contrasts with the logic model in terms of having an adult to support all student learning whilst using the programme.

Control group activity

As the evaluation was an in-school design, it was important to consider what provision was normally made for struggling readers and if the programme resulted in any changes to this provision for the pupils in the control condition. It was also important to know whether children participating in the Lexia programme received just Lexia, or benefitted from usual provision and the additional supplement, as well as what was lost as part of participating in the intervention.

The pre-intervention survey asked teaching staff to describe their current reading provision, both in general and for struggling readers. As the question was asked in free text format to gather as much detail as possible, the data also reflects what the respondents chose to emphasise and is not necessarily a full picture of the provision in each school.

General reading provision prior to the intervention

The question of reading practice in general was addressed by 135 respondents from the 57 participating schools. All schools applied a mix of individual, group, and class-based reading provision and, whilst differences of emphasis within the mix and the use of different materials were evident, common practices across the schools emerged and were coded as seen in Table 36.

Table 36: Reading provision prior to implementation

Reading feature listed	Schools N (%)
Guided reading	50 (88)
Guided and reciprocal reading	4 (7)
Reciprocal reading (guided reading not mentioned)	1 (2)
One to one reading with an adult	45 (79)
One to one reading and independent reading	6 (11)
Independent reading (one to one not mentioned)	3 (5)
Phonics (Read Write Inc.)	42 (74)
Comprehension skills activities	37 (65)

n = 57 schools.

Pre-intervention provision for struggling readers

Most schools (n = 49, 86%) reported delivering some form of reading provision on a daily basis although this frequency was often referenced in relation to struggling readers: the provision for the more able readers appeared more typically to be three or four times a week. Also, for struggling readers, one to one reading sessions were more frequently delivered than they were for the more able readers for whom class or group-based provision was the usual practice.

General reading provision during the intervention

Post-test survey respondents were asked whether the non-Lexia pupils were doing reading lessons or some other lesson or activity while the Lexia pupils were receiving the Lexia programme. Table 37 shows the breakdown of responses by these broad categories (82% of schools responded). Where additional detail was given, the main lessons, other than a reading lesson, were described as ‘topic’ (eight schools) or ‘foundation’ (six schools), and the main activity other than lessons was ‘assembly’ (22 schools).

Table 37: Control pupil activities during Lexia sessions

Reading lessons	Other lessons	Other activity (e.g. assembly)	School n (%)
✓			5 (10)
	✓		10 (19)
		✓	6 (12)
✓	✓		2 (4)
	✓	✓	7 (13)
✓	✓	✓	12 (23)
Missing: n = 9; n/a—sessions delivered before the start of school: n = 1			10 (19)

n = 52 schools.

Of the ten schools represented by the interviews, three scheduled the Lexia sessions alongside assembly and the remaining seven scheduled them during lesson time, but with varying attempts to avoid taking children out of ‘core’ lesson activity. Hence, in some schools, the non-Lexia children were doing other literacy-based activities and in others, they were variously engaged in registration, settling activities such as mindfulness, preparation for subsequent lessons such as PE, finishing activities, or looking at homework.

‘[The non-Lexia children] are doing ... guided reading ... dictionary work or practising their spellings or a bit of handwriting ... on a rotation kind of basis’ (teacher).

‘It’s a mix of things, but it’s often getting [the non-Lexia children] settled [with] some mindfulness exercises. Sometimes it’ll be handwriting, or they could just be getting ready for PE’ (Year 2 teacher).

‘It tends to be the wrap-up part of the day, so [the non-Lexia children] might be finishing off work they’ve done elsewhere in class, or it might be a story or something like that’ (TA).

‘[The non-Lexia children] are doing RE, history, geography—topic things usually—science ...’ (Year 2 teacher).

‘It’s the last half hour of the day so [the non-Lexia children] might be finishing off the foundation work they started, doing whole-class guided reading, showing and discussing their homework—just finishing off whatever we were doing that day’ (Year 2 teacher).

Overall, from both the survey and interview data, it appears that Lexia provision was in addition to, rather than instead of, normal supplementary reading activities in approximately two-thirds of participating schools.

Provision for struggling readers during the intervention

Where phonics continued to be provided during the intervention period, this was usually because the struggling readers were still working to complete a Year 1 programme such as Read Write Inc. (For Year 2 pupils who did not ‘pass’ their Year 1 national phonics assessment, schools are required to provide additional phonics provision in Year 2 to enable them to meet the specifications by the end of Year 2.) Additional one to one reading sessions were also frequently provided to support struggling readers, though the nature of the one to one provision varied considerably as it was delivered variously by teachers or TAs, adult relatives, or other pupils in a higher ability group. This one to one provision, as usually delivered, was somewhat constrained or occasionally unavailable during the intervention in a small number of schools where finite resources were concentrated on supporting the Lexia pupils.

Outcomes

Outcomes available from the survey data covered the following:

- programme deliverers' experience of using Lexia;
- programme deliverers' perception of the pupil experience of using Lexia;
- suitability of Lexia for different groups of pupils;
- respondent views on the least/most effective aspects of the Lexia programme; and
- whether respondents would recommend the Lexia programme to other schools.

Programme deliverers' experience of using Lexia

As seen in Table 38, more than 75% of respondents 'strongly' or 'somewhat' agreed with the following statements:

- the Lexia programme was easy to access (n = 47, 89%);
- there was sufficient space and equipment available in the school to deliver the programme (n = 40, 76%);
- the reports were useful in tracking progress (n = 43, 81%); and
- the support received from LexiaUK was satisfactory (n = 41, 78%).

Just over 50% of respondents strongly or somewhat agreed that it was easy to fit the delivery of the programme into the school day (n = 33, 63%) and for the facilitator to find the time required to deliver the programme (n = 28, 53%). The statement that respondents agreed with slightly less (n = 26, 49%) concerned the usefulness of paper-based resources although, as indicated in the Implementation and Fidelity section above, this may have been as much to do with the constraints on time and staff resources as any objective rejection of their potential value.

Table 38: Experience of using Lexia

	Strongly agree n (%)	Somewhat agree n (%)	Neither Agree nor disagree n (%)	Somewhat disagree n (%)	Disagree n (%)	Not applicable n (%)
Easy to access the Lexia program	26 (49)	21 (40)	2 (4)	2 (4)	0 (0)	2 (4)
Easy to fit Lexia delivery into the school day	12 (23)	21 (40)	7 (13)	7 (13)	6 (11)	0 (0)
Sufficient space available in the school to deliver the programme	22 (42)	18 (34)	4 (8)	5 (9)	3 (6)	1 (2)
Sufficient equipment in the school to deliver the programme	26 (49)	21 (40)	1 (2)	4 (8)	0 (0)	1 (2)
Easy to find time for the facilitator to deliver the programme	12 (23)	16 (30)	10 (19)	7 (13)	7 (13)	1 (2)
The paper-based resources were useful	6 (11)	20 (38)	15 (28)	5 (9)	3 (6)	4 (8)
The reports were useful in tracking pupil progress	20 (38)	23 (43)	6 (11)	1 (2)	2 (4)	1 (2)
I was satisfied with the support I received from LexiaUK	21 (40)	20 (38)	8 (15)	2 (4)	1 (2)	1 (2)

Pupils' experience of using Lexia

As seen in Table 39, between 78% and 91% of survey respondents 'strongly' or 'somewhat' agreed with the statements that pupils found the Lexia programme engaging, easy to use, were able to work independently during the sessions, and benefited from using it. Around two-thirds felt that pupils enjoyed using Lexia and were able to apply their learning from Lexia in other lessons. There was less agreement about the degree of scaffolding that pupils required from the facilitator, with 44% 'strongly' or 'somewhat' agreeing that they did compared with 57% recording a neutral or negative response. As seen from the observations above, much of this scaffolding tended to be TA focus on a particular child or children during the sessions.

Table 39: Survey respondents' reports of pupil experiences of the programme

	Strongly agree n (%)	Somewhat agree n (%)	Neither Agree nor disagree n (%)	Somewhat disagree n (%)	Disagree n (%)	Missing n (%)
Pupils enjoyed using Lexia	21 (40)	22 (42)	4 (8)	4 (8)	1 (2)	1 (2)
Pupils were engaged during the Lexia sessions	19 (36)	22 (42)	6 (11)	6 (11)	0 (0)	0 (0)
Pupils found it easy to use Lexia	21 (40)	27 (51)	4 (8)	1 (2)	0 (0)	0 (0)
Pupils worked independently during Lexia sessions	21 (40)	25 (47)	6 (12)	0 (0)	1 (2)	0 (0)
Pupils were able to apply their learning from Lexia in their other lessons	8 (15)	26 (49)	12 (23)	5 (9)	1 (2)	1 (2)
Pupils needed a lot of scaffolding from the facilitator during Lexia lessons	3 (6)	20 (38)	11 (21)	13 (25)	5 (9)	1 (2)
I feel that pupils benefited from using Lexia	20 (38)	21 (40)	9 (17)	2 (4)	1 (2)	0 (0)

Pupil experience of the programme

In the focus groups, pupils themselves were also asked about their general experience of the programme with particular attention to their:

- confidence in using the programme;
- enjoyment of the programme; and
- perceptions of learning.

Three pupils across two of the focus groups mentioned that the programme had improved their confidence:

'I like Lexia because we have harder things, and we can do it ourselves ... we think in our head we can do it' (pupil).

Linked to this was pupils' awareness and reliance on scaffolding by the Lexia facilitator. Four out of the five focus groups, when asked, mentioned scaffolding by the Lexia facilitator. One focus group, however, mentioned that the teaching assistant generally was occupied during this period listening to other children read or lesson planning:

'She sometimes writes in other people's books when she's supposed to, and then she reads with people' (pupil).

This meant that:

'I don't know how to do this, "[TA] said this is a reading practice, I can't tell you, you need to do it yourself.' (pupil).

Overall the children tended to report enjoying using the programme. This was in part due to a lack of access to IT facilities within school—‘because we don’t go on them often’ (pupil)—but also due to features of the programme itself, including certificates, moving across levels and thereby measuring progress (‘winning’), and themes associated with particular levels. They also discussed the repetition aspect of the software, that is, on completion they may have to repeat a module and this form of consolidation was valued:

‘You have to do it twice to know if you’re not really sure if you’ve done it good’ (pupil).

All pupils who discussed their learning as a result of the programme agreed that Lexia was helping them to learn. Seven pupils in focus groups said it was teaching them ‘to read’, five that it was teaching them ‘words’, three each mentioned ‘sounds’, ‘spelling’, and ‘writing’; some stated more than one aspect. However, across all five focus groups, the children were keen to discuss the levels they had passed and showed high levels of awareness of where they were in the process of the programme. This indicated their personal progression, learning and a growth in confidence, although there was a tendency to compare within each group:

‘I’m on level five.’

‘I’m on level six.’

‘I’m on level seven’ (pupils).

The certificates were also appreciated as marking progression although these were particularly mentioned in those schools where certificates were handed out publicly within the school as opposed to individually.

Perceptions of programme effectiveness

The post-test survey asked respondents to identify, in free-text format, aspects of the programme that they thought worked least, or most, effectively.

Twenty-four respondents (from 21 schools) who thought the programme was very effective said there was nothing they could think of to improve it. A further 45 respondents (from 36 schools) identified a variety of elements for potential improvement, of which those highlighted in Table 40 were the most commonly mentioned.

Table 40: Aspects highlighted for improvement by multiple respondents

	n
Too ‘Americanised’—partly in pronunciation, partly in terminology	10
Repetition could be discouraging	9
Difficult to fit into the schedule, too long	6
Unsuitability of programme for some children	4
Phonics element	3

n = 45.

The comments below are illustrative:

‘The programme had many Americanisms—with references to baseball, raccoons, etc.—and did not follow the U.K. curriculum closely enough’ (dyslexia specialist teacher).

‘[The children] struggled to understand the accent of the computer sometimes’ (assistant head).

‘When children are on green level and have almost completed the bar and then make a mistake, they get very disheartened about starting the whole bar again’ (Year 2 teacher).

Most effective aspects

Seventy-five respondents (77%) from across 50 schools answered the survey question about perceptions of the most effective aspects of the programme. Apart from eight respondents (8%) who said they thought ‘all of it’ was effective, the aspects identified by the rest were extremely diverse. Sixteen (16%) reported those aspects that they thought the children liked best, which ranged from the animations and variety of activities through to the rewards and certificates they obtained by completing the different levels. These respondents also referred to the children’s enjoyment of the computer- and game-based design of the programme, both of which made the learning more fun.

‘The children were highly motivated by the certificates for completing levels and adored the animations for completing a lesson’ (TA).

‘The children saw it as if they were learning through a game therefore would be much happier to spend time on the programme’ (TA).

‘The tablets made the children feel they were doing something other than reading’ (Year 2 teacher).

Among the comments from the 61 respondents who volunteered some detail about what they had found particularly effective, a few common themes emerged, as shown in Table 41.

Table 41: Effective aspects identified by multiple respondents

	N
Role of rewards (e.g. certificates)	11
The visuals	8
Encouragement of independent working (reading)	8
The game-based design	7
Engaging content	6
Structure—completion of, and automatic adjustment to, achievable levels	6
Computer-based	4
Use of repetition	4

n = 61.

The comments below illustrate some of these themes:

‘The way the levels were structured and the way the children could see the progress they had made was very effective as was the award of the certificates’ (TA).

‘The visual aspect was good. Particularly for story sequencing’ (HLTA).

‘It was effective in allowing them to independently work through problems by building up resilience’ (TA).

‘It is a very well designed app that meets the individual needs of the children. It uses ‘over learning’, which is what’s needed but does this effectively by using different activities/games’ (Year 2 Teacher).

‘The children were engaged in the lessons and loved it when they finished a section. They were very focused and enjoyed the control that they had. They became more confident in their learning and the progress was evident in their class work’ (HLTA).

Other effective aspects identified in isolated comments included:

- interactive;
- use of reports;

- phonics content;
- application to problem solving; and
- design and layout.

Finally, by way of an overall measure of endorsement, respondents were asked if they would recommend the programme to other schools. As Table 42 shows, nearly two thirds said yes, and a further third would consider it. Only three were unequivocal in not being able to recommend it.

Table 42: Would you recommend Lexia to other schools?

	n (%)
No	3 (3)
Maybe	35 (36)
Yes	60 (61)

n = 98.

Cost

The evaluation team collected cost data from LexiaUK and from the participating schools with the intent of establishing the average costs of implementation of Lexia Core5 Reading (via survey and interview). We estimate average marginal costs per pupil per year for schools implementing the Lexia programme. Given the date that this evaluation was commissioned, we have used the 2016 cost guidance (EEF, 2016), bearing in mind the updated 2019 guidance where possible (EEF, 2019). Whereas usual practice is to estimate costs over a three-year period, the package offered to schools by the EEF covered two years and so our main costings are based on an estimate of the two-year packages used during the evaluation and a three-year licence, which would reflect whole-school conditions. In addition, in this trial schools were expected to deliver the programme to six or seven pupils, on average, whereas normal, whole-school delivery practice would be to deliver to approximately 15 pupils at a time.¹² Consequently, we provide costings for both scenarios.

The main financial costs of implementation were:

- the Lexia programme package;
- IT equipment (including headphones); and
- printing of paper-based resources.

Additional time costs were:

- staff training; and
- TA delivery time.

Financial costs

The Lexia programme package included 30 named student licences (although licences can be reallocated over the licence period). The package also included three one-hour training sessions for school staff and ongoing support. For the evaluation, the schools contributed £500 excluding VAT. The total cost of a two-year package (outside of the evaluation) was £2,460 excluding VAT: assuming 20% VAT this would be £2,952, although schools can claim the VAT back so it is not included in the costings. The total cost of 30 licences for a three-year package is estimated at £3,690. We understand that the purchase price may be significantly less than estimated due to discounting structures applied by LexiaUK on multi-year and larger volume license purchases.

¹² Correspondence with LexiaUK.

In the post-evaluation survey and during the interviews, schools were asked about additional costs. Four of the ten interview schools indicated that there were no additional costs involved in using the programme, outside staff time (see below). Five of the ten interview schools mentioned printing paper-based resources although most (four of the five) felt it was minimal and a small amount compared to the general printing conducted by the school. None of the respondents to the survey mentioned printing costs. Consequently, printing has not been included in the overall costings. The reports can be viewed online as well as printed, and given the interview responses generally did not mention printing reports (only one out of ten) we have assumed reports were viewed online. Thirty-two schools of the 52 surveyed and two of the ten interview schools mentioned having to purchase headphones for the programme. Schools interested in taking up the programme should be made aware of this requirement as they are not always commonly used in schools. Eleven schools in the survey mentioned purchasing additional IT equipment (laptops, ipads, or PCs). This was a small number. We have, however, included them in the costs using an estimate for a medium priced tablet although we believe this is an overestimate given the proportion of schools reporting the need to purchase this equipment. Consequently, headphone and IT equipment are included as prerequisite costs in Table 43 below for two-year licences (the evaluation conditions). Table 44 provides the same information for the three-year licences (that is, whole-school delivery). It is also important to note that this equipment can be used for other purposes so in that sense this costing is higher than would actually be the case for most schools.

Table 43: Prerequisite costs for two-year licences (evaluation conditions)

Item	Type of cost	Cost	Total cost over 2 years	Total cost per pupil per year over 2 years*
Headphones	Prerequisite cost per school	£13 per pupil x 7 = £91	£91	£1.52
IT equipment (PC, tablets, laptops)	Cost per school	£350 per pupil x 7 = £2,450	£2,450	£40.83
Total	-	-	£2,541	£42.35

* Assuming 60 pupils over two years.

Table 44: Prerequisite costs for three-year licences (whole-school delivery).

Item	Type of cost	Cost	Total cost over 3 years	Total cost per pupil per year over 3 years*
Headphones	Start-up cost per school	£13 per pupil x 15 = £195	£195	£2.17
IT equipment (PC, tablets, laptops)	Start-up cost per school	£350 per pupil x 15 = £5,250	£5,250	£58.33
Total	-	-	£5,445	£60.50

* Assuming 90 pupils over three years.

Time costs

There were three one-hour training sessions offered to each school. Typically, three people attended each session although these sessions can accommodate up to six participants. There were reported to be minimal additional support requests during the implementation period so these have not been included in our estimates. Additional training is also

included in the second year of the package through LexiaUK reaching out to all customer annually and with schools having access to live and recorded training webinars. However, as we have no data on this it has also not been included in the costings. Given that the training takes place online it would only incur three hours per staff member (up to six staff members). In the survey, only one school mentioned having to increase TA employment hours to cover delivering the Lexia programme (by 15 minutes a day). All other schools that mentioned staff time instead highlighted that this involved redeployment from other areas. Consequently, this has not been costed for.

Time costs for training are included in Table 45 and those for personnel preparation and delivery under evaluation conditions in Table 46. Please note that as the training involved twilight sessions there were no staff cover costs. These only cover the time, preparation, and delivery costs for the Lexia evaluation children.

Table 45: Total time devoted by school staff for training under evaluation conditions

		Year 1		Year 2*	
		Number of school staff	Approximate number of hours	Number of school staff	Approximate number of hours
Training	Member of SMT	1	3	-	-
	Year 2 teacher	1	3	-	-
	TA	1	3	-	-

* We do not have figures for attendance at training in the second year of the licence.

Table 46: Total time devoted by school staff for preparation and delivery under evaluation conditions

		Year 1		Year 2	
		Number of school staff	Approximate number of hours per week	Number of school staff	Approximate number of hours per week
Preparation*	Year 2 teacher	1	0.50	1	0.50
	TA	1	1	1	1
Delivery	Year 2 teacher	1	See note **	1	-
	TA	1	2	1	2

* This includes checking the reports and accessing the paper-based resources. The majority of teachers (63%) reported checking the reports once a week but the paper resources was predominantly when a child was indicated to need them by the software (31%) or weekly (21%). Consequently, the teacher time is a very rough estimate. In only 5 out of 52 schools did SMT members indicate that they had accessed the reports and as we are unsure of the frequency of this access they have not been included here.

** In only six of the 52 schools did Year 2 teachers state they delivered the programme and in four of these it was alongside the TA, which means the amount of delivery they were involved in and the time costs are uncertain.

We have not included time costs for a three-year licence for whole-school delivery as this would depend upon the number of teachers trained and the number of session run within the school although estimates can be made based on the above tables (ie. if delivered to 15 children across two groups per week number of TA hours would be approximately 10 per week).

Number of students

In this evaluation there were approximately six to seven Year 2 pupils in the intervention group in each school. However, the package offered allowed schools to have up to 30 pupils undertaking the programme at any one time, with the opportunity to move licences across pupils over the two-year period. As seen above, in one observation group, 22 pupils were included in one session across year groups (Years 2 to 6), which would be more cost effective than the evaluation conditions which only included approximately six pupils with one TA. In addition, LexiaUK indicated that a typical group size for struggling readers would be 15 with one TA although it would depend upon school needs (that is, whether a whole-school approach was being adopted, the number of devices available, and so on). We estimate that under whole-school delivery conditions this would result in two delivery groups four days a week per week per school (that is, four hours TA time per week plus preparation time of approximately two hours per week). We assume that existing TA resources would be diverted to the intervention rather than new members of staff employed as this was indicated in our

interviews, although one school indicated that the TA had an additional one hour per week of preparation time (checking reports, printing out certificates and resources, and so forth).

As indicated above, during the evaluation there were only six to seven Year 2 pupils included in the intervention. However, outside of these pupils, the additional licences included in the overall package could be transferred between pupils once they were deemed by the school to have satisfactorily reached a desired standard. LexiaUK estimated approximately 60 pupils using the programme over a two-year period and 90 pupils over a three-year period and our costs are based on this.

Table 47 provides the initial cost of the programme, as evaluated in this trial (that is, delivery to six to seven pupils at a time) including the cost for a two-year period and delivery to 60 pupils over a two-year time period. We would reiterate that this does not include any additional costs such as printing (due to schools indicating this was minimal) and any additional IT equipment needed (which, as noted, is considered a prerequisite cost). It also does not include TA delivery time given this was usually diverted from current activities (see Time Costs above).

Table 47: Cost of delivering Lexia under evaluation conditions

Item	Type of cost	Cost	Total cost over 2 years	Total cost per pupil per year over 2 years
Two-year package	Start-up cost per school	£2,460	£2,460	£41.00
Printing	Running cost per school	-	-	
Total	-		£2,460	£41.00

Table 48 and Table 49 indicate the cost per pupil assuming the package covered a three-year period, under whole school conditions, and covered 90 pupils over three years as suggested by Lexia. It should also be reiterated that these costs do not include staff time under the assumption that TAs are used to deliver the programme who would have been utilised to deliver other provision for struggling readers. Consequently, we would rate this programme as very low cost.

Table 48: Estimated cost of delivering Lexia under whole-school delivery conditions

Item	Type of cost	Cost	Total cost over 3 years	Total cost per pupil per year over 3 years
Three-year package	Start-up cost per school	£3,690	£3,690	£41.00
Printing	Running cost per school	-	-	-
Total	-		£3,690	£41.00

Table 49: Cumulative costs of Lexia—assuming delivery over three years

	Year 1	Year 2	Year 3
Under whole-school conditions (i.e. 15 pupils per group)	£3,690	-	-

Conclusion

Table 50: Key conclusions

Key conclusions
1. Children offered Lexia made, on average, the equivalent of two additional months' progress in reading compared to other children. This result has a high security rating.
2. Exploratory analysis suggests that children offered Lexia made the equivalent of two additional months' progress in word recognition and decoding skills and one additional month of progress in reading fluency and comprehension skills, on average, compared to other children. Children offered Lexia made, on average, no additional progress in Key Stage 1 national test reading raw scores compared to other children.
3. Children eligible for Free School Meals (FSM) who were offered Lexia made, on average, the equivalent of three additional months' progress in reading compared to other children eligible for FSM. However, this result has high statistical uncertainty.
4. Implementation fidelity was high. Most schools incorporated Lexia into their routine so that pupils received the normal provision for struggling readers as well as this intervention. Over three-quarters of survey respondents were satisfied with the access to the online activities, support provided, and the usefulness of the reports.
5. Most pupils managed to work independently with little scaffolding from teaching assistants. However, teachers and teaching assistants felt that Lexia was unsuitable for a small number of pupils, especially those with Special Educational Needs and Disabilities and those whose reading level was very low.

Overall, this evaluation supported the logic model, as agreed with LexiaUK. There is evidence that participation in the programme resulted in approximately two months' additional progress for children in the intervention group compared to those in the control group in overall reading (WRMT-III composite score). This was particularly reflected in the Word Identification and Word Attack subscales (both two months' progress), which are the key building blocks for this age group in terms of literacy development (focusing on key phonics and word recognition skills) but also with one month's additional progress in comprehension and fluency skills, which are important for further literacy development. There was no observed additional month's progress in the KS1 raw reading scores. However, none of the results were statistically significant. The trial recruited to target in terms of the number of schools and the observed correlation between baseline and follow-up WRMT-III composite score was higher than anticipated, which meant the trial was ultimately well powered to detect an estimated effect size of at least 0.19 of a standard deviation. We had originally powered the trial for a minimum detectable effect size of 0.2, however, the observed effect size was smaller than 0.19 and so the primary result was unable to reach statistical significance. Observed effect sizes, whether statistically significant or not, need to be interpreted in terms of whether the difference would be deemed meaningful in practice.

The interaction between intervention and FSM was statistically significant at the 10% significance level suggesting that there was evidence that the intervention effect for FSM pupils is different to that for non-FSM pupils. The subgroup analysis using FSM status demonstrated a statistically significant difference between intervention and control arms, representing an effect of 3 months' additional progress, which is greater than the effect shown in the primary analysis. These results suggest that the Lexia programme may be more effective in FSM pupils; however, we have a limited sample on which to base this finding and the study was not powered to detect an effect in any subgroup.

Training and support appeared to be strong and appreciated by schools. However, despite a member of the SMT in each school attending training there was no strong evidence that this participation translated to wider school and teacher awareness of phonics, fluency, and comprehension practices given that the programme was mainly delivered by TAs with some Year 2 teacher support and collaboration. In addition, access to reports was mainly focused on TAs and Year 2 teachers. Consequently, we have altered the logic model to reflect this (see Figure 4) because there was no evidence of a change in the wider school ethos (as expected from the original logic model, Figure 1, which posited change in teacher and school phonics and comprehension awareness and practices). However, as researchers we are aware that our focus was on Year 2 struggling readers and there was wider delivery within the school and, under different conditions, this may have been explored further and would therefore suggest that any further studies explore a wider, whole-school delivery model either with struggling readers or with the school population as a whole. Given the impact on fluency, however, we have included this in the revised logic model (Figure 4).

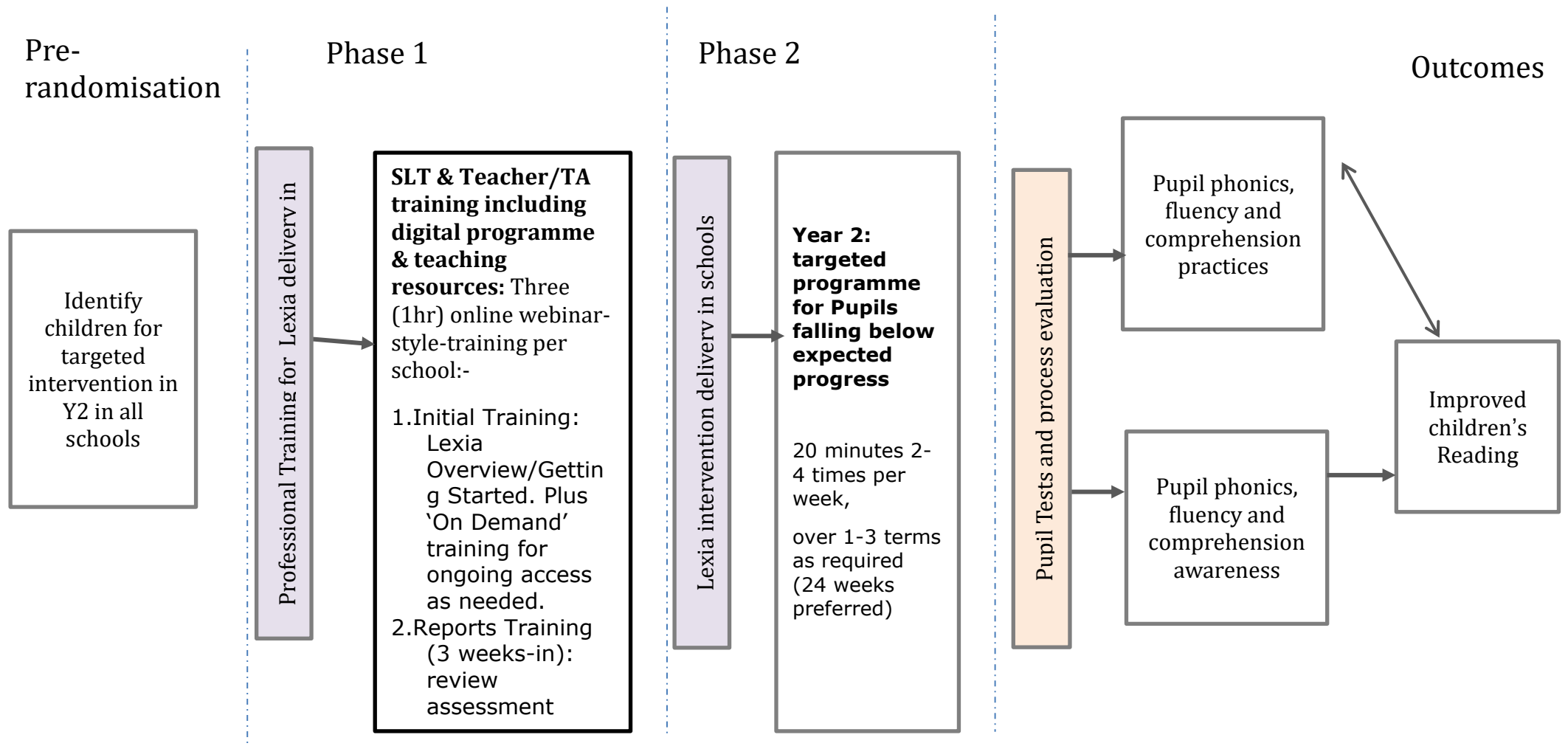
Interpretation

The impact evaluation suggests there may be some benefit of using the programme—one to two months of additional progress for intervention children compared to those in the control group for Year 2 struggling readers. This supports the existing evidence for the programme (see Introduction and, for example, Brooks, 2016) and, with some focus on phonics, use of the programme for struggling readers in Year 2 fits with U.K. policy to require pupils who do not attain requested levels in the Year 1 phonics check to repeat the check at the end of Year 2. Whilst there were some issues with translation to a U.K. context, the evidence from schools does not suggest these were large (mentioned by less than a sixth of respondents). There were few issues relating to access to the programme and, in the cases where they did occur, the short and efficient set-up times by schools alongside the responsiveness of the LexiaUK support team suggests these were reduced and managed as much as was possible. However, compliance, using programme data, was high and does appear to correlate with the results. Whilst there were no statistically significant findings relating to compliance data they do appear to be in line with what would have been expected. In particular, higher dosage would result in higher overall attainment. However, given the adaptability of the software it would be expected that already-higher-achieving pupils (within the struggling reader definition used in this study, which was variable between schools) may not have needed to make as many gains in order to meet the expected level for their year group (although this will be clearer once the KS1 data is made available and analysed). Overall, we feel the results of this study are reassuring in terms of schools being able to integrate the programme within their wider school curriculum. Pupils also enjoyed the programme, were engaged, and suggested they felt more confident as a result of taking part.

However, the programme was primarily delivered by TAs (although this is predicted and acceptable within the programme design (see TiDiER, Table) with limited wider school involvement apart from facilitating the programme and making time and resources available. In many cases these resources were diverted from other uses. There was also minimal use of the paper-based resources (hence minimal printing costs) and accessing the reports to monitor progress—although these were accessed by the majority of schools it was unclear how they used this information—which does not fit with the programme logic model or the TiDiER table. These are areas that could be strengthened within the programme. Given that delivery is primarily through TAs they could perhaps be given more training in interpreting and communicating the reports and more time could be provided in order for them to do so. Although additional reports sessions and regular invitations to webinars are provided, there does not appear to have been much take up of these.

It is also important to consider that there is a fine balance between groups and TA delivery, that is, in this evaluation most groups were six or seven pupils in Year 2 (as seen in nine out of ten of the observations). Our cost analysis includes the costs related to the evaluation pupils and those on a whole-school, three-year delivery model are based on LexiaUK's estimates of a usual group of 15 pupils, across year groups, per session. Whilst this would reduce the delivery time and redirection of TA resources it would have implications for IT provision. These would need to be balanced and schools should be aware of this before any proposed delivery. It also means that for this evaluation children probably had more adult input than they would have had in a larger group although in many cases it was evident that some children received more input than others so we do not feel this would be an impediment for widening delivery to a more extensive range of year groups. This is an area that would be of interest for further research. Of particular concern, as discussed above, is that control pupils did not appear to receive the same level of support as they would usually as resources were diverted or intervention pupils received usual provision and the Lexia programme. In addition, it was felt by some schools that there were some pupils for whom the programme was not appropriate, particularly those with SEND for whom use of an IT programme was felt to need particularly high levels of scaffolding. Excluding this group from implementation would, in our view, mitigate some of the comments made above, that is, in terms of TA time and resources being diverted from the group as a whole and reducing one to one TA time with some pupils. We would suggest that either the programme was developed further to need less scaffolding for these pupils or that schools, by using less TA time on the programme (by having larger group implementation), would be able to provide more one to one provision for these pupils.

Figure 4: Adjusted logic model for evaluation of Lexia Reading



Limitations

There were limitations to this evaluation, in particular the focus on Year 2 struggling readers: under normal conditions, the delivery package could span across year groups and across more than six or seven pupils per year, which the research team were unable to capture. If a wider age range had been included this may have captured more whole-school ethos change and awareness as described in the original logic model (Figure 1). However, with this being an efficacy trial, there were limitations to what could be achieved in terms of scale. One particular limitation was the need to have a single year group in order to have comparable results. However, this does mean we do feel secure in our findings and were able to assess the impact of the programme on Year 2 struggling readers.

Overall, there were high levels of retention of pupils and schools, which means we can be secure in our findings under the trial conditions. Compliance, as agreed and measured, was also high, which means the programme was easily accessible and implementable by schools and the results, therefore, generalisable more widely to schools with Year 2 struggling readers. There was also no evidence of contamination given that the programme licences were for a particular child and paper-based resources were linked to that child's progress. The main issue for schools emerging from the findings are the possible impact of diverting already limited TA resources to implementing the programme. This means that we do have concerns that control pupils did not receive their normal level of provision, which may have impacted on our findings by causing bias in favour of the intervention pupils. The primary implementation was, however, by TAs and this may have impacted on the limited use of paper-based resources and use of the reports. We would suggest that the training focuses on TAs as programme deliverers and provides additional specifications for teachers on their role within the programme to ensure a wider-school focus.

Future research and publications

In terms of future research, the results suggest that a larger-scale effectiveness trial may be worthwhile, although the target sample size to detect an effect size of 0.1 would be large and therefore increase the costs of the trial. It would also be useful to assess the results from a wider, whole-school implementation or of different models of the programme—for example, across year groups as opposed to delivery to one year group (in this case Year 2 struggling readers). This would involve a two-arm, between-school trial rather than within-school as was used in this case but would be better able to assess whole-school change as posited in the original logic model. It would also be interesting to see results of using the Lexia Core Reading programme instead of, as well as in addition to, current provision for struggling readers, especially for older year groups when comprehension skills become more important. This trial could also be used to evaluate whether the greater benefit of Lexia among FSM pupils can be replicated. Finally, given the pandemic, it may be useful to consider and rigorously evaluate the home delivery mode of the Lexia programme.

References

- Azur, M. J., Stuart, E. A., Frangakis, C. and Leaf, P. J. (2011) 'Multiple Imputation by Chained Equations: What is It and How Does It Work?', *International Journal of Methods in Psychiatric Research*. DOI: 10.1002/mpr.329
- Brooks, G. (2016) 'What Works for Children and Young People with Literacy Difficulties? The Effectiveness of Intervention Schemes', Sheffield: DyslexiaSpLD Trust.
<http://www.dyslexiaaction.org.uk/files/dyslexiaaction/what-works-5th-edition.pdf>
- Castro, J. and Robnolt, V. (2016) 'Woodcock Johnson Reading Mastery: Test Review', *READ 600* (8 March 2016):
<https://rampages.us/castrojf/wp-content/uploads/sites/23866/2017/04/Reading-Mastery-Paper.docx>
- Chambers, B., Slavin, R. E., Madden, N. A., Abrami, P. C., Tucker, B. J., Cheung, A. and Gifford, R. (2008) 'Technology Infusion in Success for All: Reading Outcomes for First Graders', *Elementary School Journal*, 109 (1), pp.1–15.
- Cheung, A. and Slavin, R. E. (2012) 'Effects of Educational Technology Applications on Reading Outcomes for Struggling Readers: A Best Evidence Synthesis', Baltimore, MD: Johns Hopkins University, Center for Research and Reform in Education (June 2012).
- Cohen, J. (1988) *Statistical Power Analysis for the Behavioral Sciences (2nd edn)*, New York: Academic Press.
- DfE (2019a) 'Schools, Pupils and Their Characteristics: January 2019', Department for Education:
<https://www.gov.uk/government/statistics/schools-pupils-and-their-characteristics-january-2019>
- DfE (2019b) 'Phonics Screening Check and Key Stage 1 Assessments in England, 2019', Department for Education:
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/851296/Phonics_screening_check_and_key_stage_1_assessments_in_England_2019.pdf
- DfE (2021) 'Special Educational Needs in England: Academic Year 2019/20', Department for Education:
<https://explore-education-statistics.service.gov.uk/find-statistics/special-educational-needs-in-england>
- Dunn, G., Maracy, M. and Tomenson, B. (2005) 'Estimating Treatment Effects from Randomized Clinical Trials with Noncompliance and Loss to Follow-Up: The Role of Instrumental Variable Methods', *Statistical Methods in Medical Research*. DOI: 10.1191/0962280205sm403oa
- EEF (2017) 'Improving Literacy in Key Stage 1: Guidance Report', London: Education Endowment Foundation:
https://educationendowmentfoundation.org.uk/public/files/Publications/Literacy/KS1_Literacy_Guidance_2017.pdf
- EEF (2018) '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) 'Cost Evaluation Guidance for EEF Evaluations', London: Education Endowment Foundation:
https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting_up_an_Evaluation/Cost_Evaluation_Guidance_2019.12.11.pdf
- Fairhurst, C. and Rex, S. (2019) 'Statistical Analysis Plan: Lexia', London: Education Endowment Foundation:
https://educationendowmentfoundation.org.uk/public/files/Projects/Lexia_SAP.pdf
- Feister, L. (2013) 'Early Warning Confirmed: A Research Update on Third-Grade Reading', Baltimore, MD: Annie E. Casey Foundation.
- Higgins, S., Katsipataki, M. and Coleman, R. (2014) 'Reading at the Transition: Interim Evidence Brief', London: Education Endowment Foundation:
https://educationendowmentfoundation.org.uk/public/files/Publications/EEF_Interim_Evidence_Brief_Reading_AtTheTransition.pdf
- Hedges, L. V. (2007) 'Effect Sizes in Cluster-Randomized Designs', *Educational and Behavioral Statistics*. DOI: 10.3102/1076998606298043
- Higgins, S., Xiao, Z. and Katsipataki, M. (2012) 'The Impact of Digital Technology on Learning: A Summary for the Education Endowment Foundation', London: Education Endowment Foundation:
[https://educationendowmentfoundation.org.uk/public/files/Publications/The_Impact_of_Digital_Technologies_on_Learning_\(2012\).pdf](https://educationendowmentfoundation.org.uk/public/files/Publications/The_Impact_of_Digital_Technologies_on_Learning_(2012).pdf)

- Hoffmann, T. C., Glasziou, P. P., Boutron I., Milne, R., Perera, R., Moher, D., et al. (2014) 'Better Reporting of Interventions: Template for Intervention Description and Replication (TIDieR) Checklist and Guide', *BMJ*, 348:g1687. DOI: 10.1136/bmj.g1687
- IBM Corp. (2019) 'IBM SPSS Statistics for Windows' (Version 26.0), Armonk, NY: IBM Corp.
- Macaruso, P., Wilkes, S., Franzén, S. and Schechter, R. (2019) 'Three-Year Longitudinal Study: Impact of a Blended Learning Program – Lexia Core5 Reading – on Reading Gains in Low-SES Kindergarteners', *Computers in the Schools*, 36 (1), pp. 2–18.
- Ness, M., Couperus, J. and Wiley, M. (2013) 'A Comparison Study of the Effectiveness of the Lexia Reading Program', *Kairaranga*, 14 (1), pp. 16–24.
- O'Callaghan, P., Mclvor, A., McVeigh, C. and Rushe, T. (2016) 'A Randomised Controlled Trial of an Early-Intervention, Computer-Based Literacy Program to Boost Phonological Skills in 4- to 6-Year Old Children', *British Journal of Educational Psychology*, 86 (4), pp. 546–558.
- Ofsted (2020) 'State-Funded Schools Inspections and Outcomes as at 31 March 2019': <https://www.gov.uk/government/statistics/state-funded-schools-inspections-and-outcomes-as-at-31-march-2019/state-funded-schools-inspections-and-outcomes-as-at-31-march-2019>
- QSR International Pty Ltd. (2018), NVivo, Version 12.
- Quartagno, M. and Carpenter, J. (2018) 'jomo: Multilevel Joint Modelling Multiple Imputation', The Comprehensive R Archive Network.
- R Development Core Team and R Core Team (2011) 'R: A Language and Environment for Statistical Computing', R Foundation for Statistical Computing, 1, p. 409. DOI: 10.1007/978-3-540-74686-7
- Snowling, M. J., Stothard, S., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., Nation, K. and Hulme, C. (2011) 'YARC: York Assessment of Reading for Comprehension: Passage Reading' (2nd edn), London: GL Assessment.
- StataCorp (2017) 'Stata Statistical Software: Release 15', College Station, TX: StataCorp LLC. DOI: 10.2307/2234838
- Taylor, M. (2018) 'Using the Lexia Reading Program to Increase NWEA MAP Reading Scores in Grades 1 to 3', American International College, Springfield, MA. <https://files.eric.ed.gov/fulltext/ED581492.pdf>
- Tracey, L., Chambers, B., Slavin, R., Hanley, P. and Cheung, A. (2014) 'Success for All in England: Results From the Third Year of a National Evaluation', *SAGE Open*, 4 (3).
- What Works Clearinghouse (2009) 'Beginning Reading: Lexia Reading', WWC Intervention Report, Institute of Education Sciences (June 2009): https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_lexia_063009.pdf
- Woodcock, R. W. (2011) *Woodcock Reading Mastery Tests (3rd edn)*, Pearson.

Appendix A: EEF cost rating

Appendix figure 1: Cost rating

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

Appendix B: Security classification of trial findings

Appendix figure 2: Security classification of trial findings

OUTCOME: WRMT-III composite reading score

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

Threats to validity	Threat to internal validity?	Comments
Threat 1: Confounding	Low	Randomisation was performed independently by the evaluation team and replicable. Imbalance in pre-test scores is very small (0.01SD in favour of the control group) for the primary outcome measure and small for secondary outcomes (< 0.05SD).
Threat 2: Concurrent Interventions	Low	No evidence of major or unmeasured changes to control teaching.
Threat 3: Experimental effects	Moderate	The IPE identified that, in order to deliver the programme, resources (TA time) were diverted from supporting control group pupils in some schools. This would increase the likelihood of observing a positive impact of Lexia compared with business as usual.
Threat 4: Implementation fidelity	Moderate	Implementation logic is specified in the logic model and fidelity is well defined. Compliance was moderate (80%). Full data on training attendance was not available for the IPE, which makes it difficult to interpret this aspect of fidelity accurately. However, available data on

		training attendance suggests that, of the survey respondents, only 34% of respondents who delivered the programme attended all three training sessions, while 28% did not attend any training sessions. Other aspects of implementation fidelity were moderate to high.
Threat 5: Missing Data	Low/Moderate	Total missing data is moderate (11%) but there is no evidence of differential attrition and results after imputation are very similar to complete case analysis.
Threat 6: Measurement of Outcomes	Low	Appropriate outcomes selected. A less favoured (dated) assessment was used at pre-test due to unavailability of the newer assessment, but this has been clearly discussed and is unlikely to have affected impact estimates. Administrators were blind to group allocation. No floor or ceiling effects were observed.
Threat 7: Selective reporting	Low	Study is registered. No evidence of selective reporting in comparison with protocol or SAP; deviations are reported appropriately.

- **Initial padlock score:** [4] Padlocks – Cluster randomised controlled trial with MDES of 0.18 and total overall attrition of 11%.
- **Reason for adjustment for threats to validity:** [0] Padlocks – two moderate threats to internal validity with direction of biases unknown.
- **Final padlock score:** initial score adjusted for threats to validity = [4] Padlocks

Appendix C: Changes since the previous evaluation¹³

Appendix table 1: Changes since the previous evaluation¹⁴

	Feature	Pilot to efficacy stage	Efficacy to effectiveness stage
Intervention	Intervention content	Describe any changes to the content.	Describe any changes to the content.
	Delivery model	Describe any changes in the delivery mechanism (e.g. from developer-led to train-the-trainers; in-person vs online; etc.).	Describe any changes in the delivery mechanism (e.g. from developer-led to train-the-trainers; in-person vs online; etc.).
	Intervention duration	Describe any changes in the duration of delivery (e.g. shortened due to the inclusion of a pre-test)	Describe any changes in the duration of delivery (e.g. shortened due to the inclusion of a pre-test)
Evaluation	Eligibility criteria	Describe any changes in the eligibility criteria for participation in the evaluation (schools, year groups, pupils etc.).	Describe any changes in the eligibility criteria for participation in the evaluation (schools, year groups, pupils etc.).
	Level of randomisation	Not applicable to pilots.	Describe any changes to the design from efficacy to effectiveness stage to the level of randomisation
	Outcomes and baseline	Not applicable to pilots.	Describe any changes to the design from efficacy to effectiveness stage in: <ul style="list-style-type: none"> ○ Outcomes ○ Baselines
	Control condition	Not applicable to pilots.	Describe any changes to the design from efficacy to effectiveness stage to the control condition

¹³ Please delete this section if it is not applicable.

¹⁴ Delete columns from the table if they are not applicable or adjust titles as relevant.

Appendix D: Effect size estimation

Appendix table 2: Effect size estimation

Outcome	Unadjusted differences in means	Adjusted differences in means (95% CI)	Intervention group		Control group		Pooled variance
			N (missing)	Variance of outcome	N (missing)	Variance of outcome	
WRMT-III composite score	2.84	3.63 (-1.34, 8.57)	310 (38)	N/A	310 (39)	N/A	1050.5
WRMT-III word identification score	1.28	1.00 (-0.57, 2.58)	318 (30)	N/A	322 (27)	N/A	107.9
WRMT-III word attack score	1.27	1.34 (-0.40, 3.07)	318 (30)	N/A	320 (29)	N/A	128.3
WRMT-III passage comprehension score	1.12	0.92 (-0.68, 2.51)	313 (35)	N/A	320 (29)	N/A	111.7
WRMT-III oral reading fluency score	0.40	0.66 (-0.85, 2.16)	314 (34)	N/A	313 (36)	N/A	95.9
KS1 raw reading Score	-0.06	-0.34 (-1.77, 1.10)	300 (48)	N/A	308 (41)	N/A	9.5

Appendix E: Missing data analysis from version 1.0 of report

As outlined in the report, the original version of this report included multiple imputation models that did not include gender or FSM status due to difficulties accessing NPD data via the ONS Secure Researcher Service during the COVID-19 pandemic. Once NPD data could be accessed, the models were run including gender and FSM status, and the results section of this report was updated. The results of the original multiple imputation analysis are presented below.

Appendix table 3: Multiple imputation on primary outcome

	Adjusted mean difference (95% CI)	Residual standard error	Hedges' g (95% CI)	p-value
Multiple imputation by chained equations	3.49 (-1.55, 8.54)	32.39	0.108 (-0.048, 0.264)	0.175
Multiple imputation by chained equations with multilevel model	3.49 (-1.50, 8.49)	32.55	0.107 (-0.046, 0.261)	0.170

This work was produced using statistical data from ONS. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates.

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>



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

<https://educationendowmentfoundation.org.uk>

 @EducEndowFoundn

 Facebook.com/EducEndowFoundn