



Education
Endowment
Foundation

**Independent evaluation of the
Lexia® Core5® Reading Programme:
A cluster randomised controlled trial**

Evaluation report

September 2026

Rebecca Mead, Katie Sykes, Anna Brian,
James Merewood, Louis Hodge, and
Elena Rosa Speciani



The Education Endowment Foundation (EEF) is an independent charity dedicated to breaking the link between family income and education achievement. We support schools, colleges, and early years settings to improve teaching and learning for 2- to 19-year-olds through better use of evidence.

We do this by:

- **Summarising evidence.** Reviewing the best available evidence on teaching and learning and presenting in an accessible way.
- **Finding new evidence.** Funding independent evaluations of programmes and approaches that aim to raise the attainment of children and young people from socio-economically disadvantaged backgrounds. Putting evidence to use.
- **Putting evidence to use.** Supporting education practitioners, as well as policymakers and other organisations, to use evidence in ways that improve teaching and learning.

We were set-up in 2011 by The Sutton Trust partnership with Impetus with a founding £125m grant from the Department for Education. In 2022, we were reendowed with an additional £137m from government, allowing us to continue our work until at least 2032.

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



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



0207 802 1653



info@eefoundation.org.uk



www.educationendowmentfoundation.org.uk



Department
for Education

What
Works
Network The logo for the What Works Network, featuring a stylized 'W' made of three overlapping shapes.

Table of contents

About the evaluator	3
Executive summary	4
Introduction	6
Methods	14
Impact evaluation results	32
Implementation and Process Evaluation results	49
Conclusion	62
References	65
Appendix A: The 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 and other tables	73
Appendix E: Outcome distributions	79
Appendix F: Diagnostic plots	82
Appendix G: Memorandum of Understanding	103
Appendix H: Parent information sheet, privacy notice, and withdrawal form	113
Appendix I: Randomisation code	121

About the evaluator

The project was independently evaluated by a team from RAND Europe.

Contact details:

Elena Rosa Speciani
RAND Europe,
Eastbrook House,
Shaftesbury Road,
Cambridge,
CB2 8DR,
United Kingdom

Email: erspeciani@randeurope.org

Acknowledgements

We would like to thank the project team, Emily Hutton, Lili Xu, and Elizabeth Kadar (RAND Europe), Lydia Lymperis, Miguel Subosa, and Emily MacLeod (formerly RAND Europe). We would also like to thank all the schools, teachers, and pupils who participated in the study, as well as the teams at LexiaUK (Rob Kay and Norman Robinson), Elkan (led by Alex Hall), and Queen's University Belfast (Dr Maria Cockerill and Joanne O'Keefe). Finally, many thanks to the Education Endowment Foundation team: Lauren Spinner; Rahul Rathore; Amy Ellis-Thompson; and Fabiola Clemente.

Executive summary

The project

Lexia® Reading Core5® (hereafter Lexia) Programme is a computer-based learning system that provides pupils in Key Stages 1 and 2 with individualised reading instruction and practice in six areas: phonological awareness; phonics; structural analysis¹; automaticity; fluency; vocabulary; and comprehension. The programme includes online personalised learning modules, real-time reporting on pupils’ progress, and paper-based resources to further support instruction.

This project aimed to evaluate the effectiveness of Lexia in improving the reading ability of struggling readers aged six to seven (Year 2) who were falling behind the expected reading standards for their age. Lexia was delivered by trained teachers and teaching assistants (TAs) for four to five 30-minute sessions per week, for a minimum of 12 weeks, to struggling readers in Year 2. Each school selected between 8 and 16 struggling readers to receive the intervention in addition to normal reading/English lessons. Pupils used the software independently, while teachers and TAs provided guidance and targeted support informed by the programme’s reporting system. Teachers and TAs received three hours of training and ongoing support from Lexia throughout the programme. The delivery team comprised of Queen’s University Belfast and LexiaUK.

The trial was conducted in 224 primary schools across England during the academic years 2023/2024 and 2024/2025. Half of the participating schools were located within an Education Investment Area (EIA). The study was funded by the Education Endowment Foundation (EEF) through the Department for Education’s Accelerator Fund to support schools in disadvantaged areas. The trial had a split-cohort design with data collected over two cohorts of Year 2 pupils across two academic years. This large-scale effectiveness trial used a two-arm clustered randomised controlled trial (RCT), complemented by a mixed-methods implementation and process evaluation (IPE). This included teacher surveys, interviews, and observations of reading sessions in schools.

Table 1: Key conclusions

Key conclusions	
1.	Pupils in Lexia schools made two additional months’ progress in reading, on average, compared to pupils in other schools. This result has a high security rating.
2.	Among pupils eligible for Free School Meals (FSM), those in Lexia schools made two additional months’ progress in reading, on average, compared to those in control schools. These results may have lower security than the overall findings because of the smaller number of pupils.
3.	Pupils in Lexia schools made two additional months’ progress in the following specific reading skills: Word Identification; Applying Phonics; and Passage Comprehension; on average, compared to pupils in control schools.
4.	Exploratory evidence suggests that increasing usage of Lexia beyond the minimum guidelines leads to greater progress in reading proficiency.
5.	Lexia was relatively easy to implement in schools which had the appropriate technology available, e.g. laptops, headphones, and internet.

EEF security rating

These findings have a high security rating. This was an effectiveness trial, which tested whether the intervention worked under everyday conditions in a large number of schools. The trial was a well-designed two-arm RCT. The trial was well-powered. Relatively few pupils (13 %) who started the trial were not included in the final analysis. The pupils who received Lexia were similar to those in the control group in terms of prior attainment. However, the trial’s security was reduced by schools in both groups delivering additional reading interventions and technology-supported instruction as standard practice. This means any improvement seen could be down to those extras rather than Lexia.

¹ See: www.lexialearning.com/sites/default/files/resources/PedagogyPaper_Core5_Structural_Analysis.pdf

Additional findings

Struggling readers using Lexia made an additional two months' progress on average, compared to similar ability pupils in control schools. This is our best estimate of impact, which has a high security rating. As with any study, there is some uncertainty around the result, and we estimate the possible impact of this programme ranges between one and three months' additional progress. We also observed two months' additional progress in the following reading skills: Word Identification; Applying Phonics; and Passage Comprehension. Lexia appears to be less effective at improving oral fluency, where we observed no additional progress.

Stronger progress was made by pupils who spent more time using Lexia, and by those who started out with higher reading ability. Pupils in Lexia schools who scored in the highest half of the baseline assessment made more progress than pupils who scored in the lowest. Pupils without special educational needs and disabilities (SEND) made more progress than pupils with SEND. This suggests that Lexia's impact was smaller for those who started out with lower reading ability compared to pupils with higher ability at baseline, and was also smaller for pupils with SEND compared to pupils without SEND. However, these findings are based on lower numbers of pupils and should be interpreted with caution. There is tentative evidence that pupils with English as an additional language (EAL) may have made more progress than pupils without EAL, but small sample sizes mean we are unable to conclude this with confidence.

There is strong evidence to support many parts of Lexia's theory of change. The impact analysis shows that Lexia led to progress in reading ability, including the skills targeted in the theory of change, such as phonics and comprehension. Most participating schools were already using technology and interventions for struggling readers. However, Lexia's personalised, technology-led approach likely enabled more struggling readers to regularly receive appropriately levelled instruction.

Most schools reported that Lexia was relatively straightforward to deliver, particularly where digital infrastructure (e.g. laptops, headphones, and internet) was already in place. Delivery was facilitated by high-quality training from LexiaUK, practitioner and headteacher commitment, and schools' allocation of equipment, space, and time.

These findings are a robust contribution to the considerable existing evidence on the effectiveness of Lexia and align closely with the earlier efficacy reports (Tracey *et al.*, 2022; Lexia®, 2024) which produced a slightly lower progress estimate (one additional month's progress) but similar positive findings on implementation.

Cost

Lexia cost an average of £4,185.63 per school (£27.91 per pupil, per year over three years), mainly consisted of the £3,850 licence. Less than half of schools reported additional purchases (e.g. headphones), although around 10% spent over an additional £1,250. Training took between 12 and 18 hours per school (four to six hours per staff member) across three years, with cover costs estimated to be £315.14 per school, per year.

Impact

Table 2: Summary of impact on primary outcome(s)

Outcome / group	Effect size (95% confidence interval)	Estimated months' progress	EEF security rating	No. of pupils	P-value	EEF cost rating
Reading proficiency (all pupils)	0.15 (0.08-0.21)	2		2,197	<0.001	£££££
Reading proficiency (pupils eligible for FSM)	0.17 (0.08-0.26)	2		724	<0.001	£££££

Introduction

Background

Evidence has shown that deficits in reading competence are cumulative: pupils who struggle to read in primary school are likely to remain struggling in later years (see Sparks *et al.*, 2014; Ferrer *et al.*, 2015; Scammacca *et al.*, 2020). These early reading skill deficits have been found to correlate with various socio-demographic factors, such as socio-economic disadvantage and use of English as an Additional Language (EAL). Primary-age pupils' reading achievement has shown a negative association with socio-economic disadvantage at the household level (Hartas, 2011). For instance, using pupil-level data generated from the Millennium Cohort Study, various studies have shown that pupils from the following types of households demonstrate lower levels of reading competence (Hartas, 2011; Lenkeita *et al.*, 2018; Russell *et al.*, 2018):

- households where parents have lower educational qualifications;
- households where parents have low-level occupations;
- households whose total income falls below the poverty line;
- single-parent households; and
- households residing in social housing tenure.

Conversely, pupils from less disadvantaged households displayed higher levels of reading attainment (Hartas, 2011; Lenkeita *et al.*, 2018; Russell *et al.*, 2018).

Language used at home is another factor that explains the reading achievement gap (Burgoyne *et al.*, 2011; Babayiğit, 2014; Lenkeita *et al.*, 2018). For example, in an evenly split sample of monolingual pupils and EAL pupils, EAL pupils displayed lower levels of reading-related skills compared to their monolingual peers (Burgoyne *et al.*, 2011). While Burgoyne and colleagues (2011) observed that EAL pupils were able to read with relatively high speed and accuracy, those same pupils also had lower levels of vocabulary knowledge, which in turn diminished their reading comprehension (Burgoyne *et al.*, 2011). Similarly, Babayiğit (2014) found that, while EAL and monolingual pupils were comparably proficient in terms of reading speed, reading accuracy, and word recognition, the latter performed better on measures of reading comprehension due to their higher level of English language proficiency. Data from the Organisation for Economic Co-operation and Development (OECD) Programme for International Student Assessment (PISA) between 2000 and 2012 showed a similar pattern, with pupils whose language spoken at home differed from the country's majority language, showing lower levels of reading achievement on average (Lenkeita *et al.*, 2018).

Having Special Educational Needs and Disabilities (SEND) is also correlated with lower levels of reading proficiency. For instance, Daniel (2024) analysed data from the UK National Pupil Database (NPD) on approximately 2.5 million pupils and found a substantial reading achievement gap between pupils with and without SEND: by Year 6, pupils with moderate, profound, or severe learning disabilities, had fallen an average of two years behind their peers in reading proficiency, while pupils with speech, language, and communication needs or specific learning difficulties were 1.5 to 1.8 years behind (although pupils with certain other forms of SEND, such as autism spectrum disorders, showed smaller proficiency gaps).

Fortunately, reading skill deficiencies in the early years can be bridged (Northrop, 2017), and early reading interventions have been effective: a meta-analysis has shown how early reading interventions yield positive outcomes for young struggling readers (Wanzek *et al.*, 2018). Among these interventions, there has been growing interest in computer-assisted instruction to help struggling readers (Kulik, 2003; Cheung and Slavin, 2012). However, the evidence regarding the effectiveness of computer-assisted reading instruction remains mixed and comes primarily from the United States (US) (Lewin *et al.*, 2019). While there is some evidence from studies conducted in the UK, this evidence remains limited, with many impact evaluations of computer-assisted reading instruction in the UK being based only on small-scale

implementations or on programmes that were delivered under optimal conditions (e.g. O’Callaghan *et al.*, 2016; Tracey *et al.*, 2022).

A previous efficacy trial (Tracey *et al.*, 2022) commissioned by the Education Endowment Foundation (EEF) to evaluate the Lexia® Reading Core5® Programme (referred to as Lexia throughout) was conducted with a sample of 697 pupils across 57 schools. Results of this efficacy trial showed significant differences in reading outcomes between pupils who received Lexia and those who did not, albeit with a modest effect size of 0.08 (one month’s learning). The observed effect size was slightly larger for pupils eligible for Free School Meals (FSM), calculated at 0.18 (two months’ learning). However, this efficacy trial was not powered to detect an effect for FSM-eligible pupils; as such, the observed subgroup effect size could not be deemed conclusive.

Our effectiveness trial provided an opportunity to explore differential effects of the programme on pupils from socio-economically disadvantaged backgrounds, supporting findings from the efficacy trial while also supporting the EEF’s broader mission to break the link between socio-economic disadvantage and educational outcomes. This focus on disadvantaged populations also aligns with the UK Government’s broader levelling-up policy (HM Government, 2022), which designated specific educationally underperforming locales as Education Investment Areas (EIAs). It also directly supports the objectives of the Department for Education’s (DfE’s) Accelerator Fund, which seeks to allocate supplementary resources to pupils who are lagging in their acquisition of English and mathematics competencies.

The present effectiveness trial was also designed to address challenges encountered during the preceding efficacy trial. For instance, while the efficacy trial randomised the sample at pupil level, the present effectiveness trial implemented randomisation at school level. This deviation was made as the efficacy trial noted that some schools diverted resources from pupils assigned to the control group (Tracey *et al.*, 2022); randomising at the school level therefore prevented the occurrence of such confounding events.

In summary, the present effectiveness trial provided a valuable opportunity to assess the impact, at a larger scale, of Lexia on young struggling readers’ outcomes—particularly given the fact that individualising reading instruction remains a common challenge for schools (Griffiths and Stuart, 2013).

Intervention

Lexia is a computer-based learning system, developed by Lexia Learning Systems LLC and made available by LexiaUK Ltd., that aims to support reading instruction at Key Stage 1 and Key Stage 2. Lexia has been implemented in over 3,000 schools in the UK and over 22,000 schools in the US. The programme provides a solution to schools’ difficulties in providing personalised support to pupils who are falling behind the expected reading standards for their age.

The programme offers the following components: i) online learning modules personalised to pupils’ learning and content mastery; ii) real-time reporting on pupils’ progress; and iii) paper-based resources to further support instruction. Further instructional support may take the form of:

- additional software use time;
- additional paper-based independent practice activities, downloadable from the software; and/or
- supplementary one-on-one or small-group face-to-face instruction based on scripted lesson plans, downloadable from the resources area of Lexia’s reporting platform, mylexia™.

Lexia targets six reading skill domains: phonological awareness; phonics; structural analysis²; automaticity; fluency; vocabulary; and comprehension. As pupils use Lexia, the software tracks their progress on the six domains and tailors

² See: www.lexialearning.com/sites/default/files/resources/PedagogyPaper_Core5_Structural_Analysis.pdf

practice and instructional activities based on their learning trajectory. While pupils are expected to use the software independently, teachers and teaching assistants (TAs) provide initial guidance on how to use the software and supplementary face-to-face instruction to reinforce learning, where needed. As such, teachers and TAs are required to attend three-hour-long online training sessions to ensure that they can deliver Lexia as intended. Schools may also request or be offered additional training sessions and ongoing support.

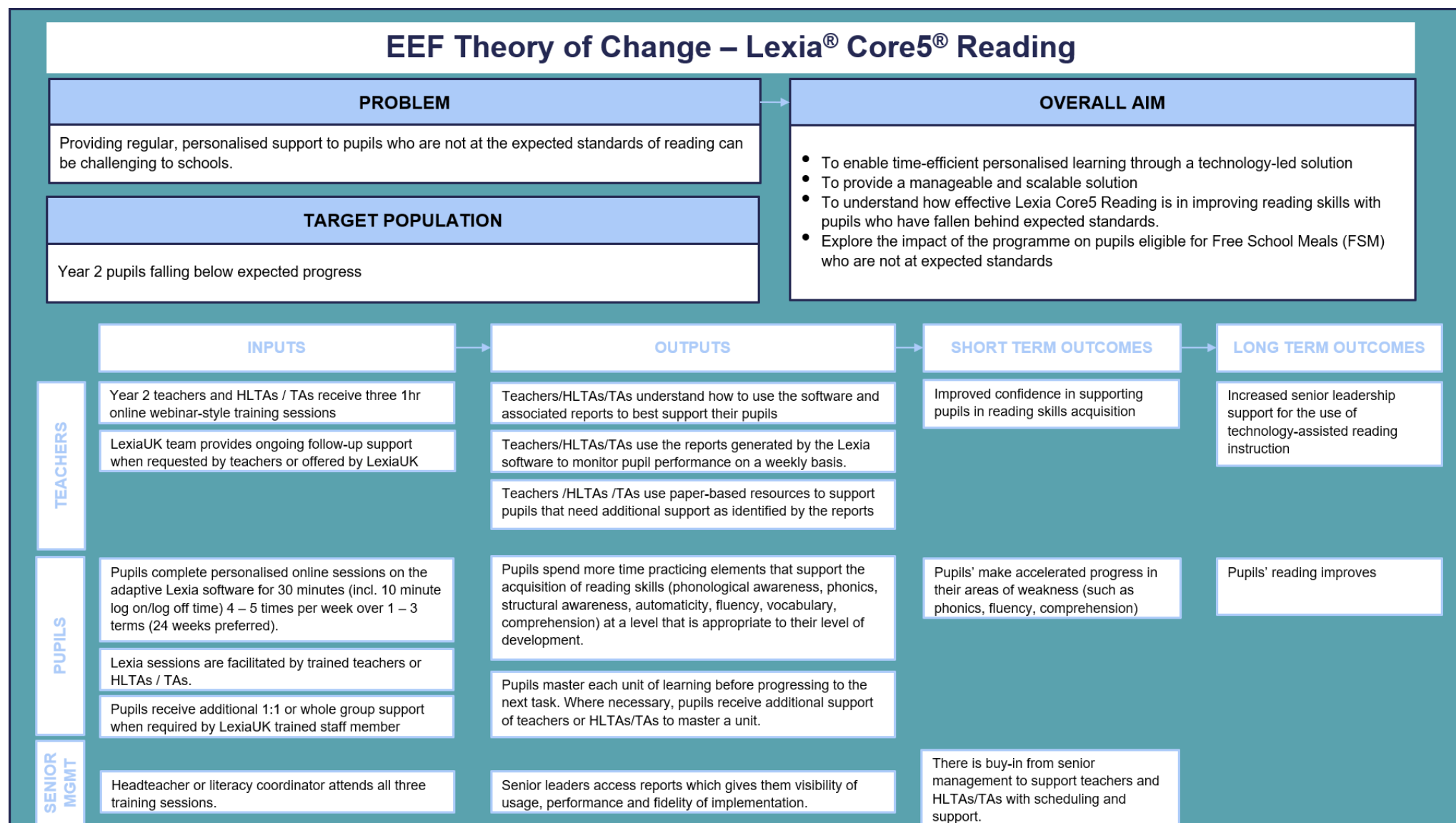
The minimum three-year licence package for Lexia allows up to 50 pupils to access the software at any one time. Lexia is commonly used in UK schools to support struggling readers, but it can be used with pupils of all reading ability levels. For the purposes of this evaluation, it was used specifically with struggling readers in Year 2, to mimic the same population used in the efficacy trial (Tracey *et al.*, 2022). When a pupil first logs into Lexia, they complete an adaptive diagnostic assessment, the result of which determines the level at which they are placed for the online learning activities. For the purposes of this evaluation, Lexia sessions were designed to typically take place four to five times per week, with each session lasting around 30 minutes, including ten minutes for pupils to log in and out of the software.

The intervention spanned two cohorts across two years, with the first delivery of Lexia taking place between September 2023 and May 2024, and the second between September 2024 and May 2025. Nominated teachers and TAs received pre-intervention training in September 2023 (for Cohort 1) and September 2024 (for Cohort 2). The split-cohort design was necessitated by constraints to delivery capacity, as explained in the 'Trial design' subsection in the 'Methods' section below.

Lexia seeks to improve readers' outcomes through the theory of change pathways laid out in Figure 1 below. As noted above, unlike many reading interventions, Lexia provides support that is tailored to individual pupils. This support is intended to enable each child to work at their own pace and level, gaining additional practice in their weaker reading skills, making accelerated progress, and ultimately becoming more capable readers. The theory of change also highlights the possible benefits of Lexia for improving teaching practices and practitioner confidence in teaching reading.

As the next section of this report demonstrates, this trial is designed to test for the presence or absence of the relevant inputs, outputs, and outcomes set out in the theory of change at both child and practitioner level.

Figure 1: Lexia® Core5® Reading theory of change



HLTA=higher level teaching assistant; MGMT=management.

Evaluation objectives

The impact evaluation seeks to answer the following primary research question (RQ):

RQ1: What is the impact of Lexia on reading proficiency, as measured by the Woodcock Reading Mastery Tests, Third Edition (WRMT-III), of struggling readers in Year 2, compared to those who do not receive Lexia?

In addition, it also responds to the following secondary RQs:

RQ2a: What is the impact of Lexia on reading and pronouncing words of struggling readers in Year 2, as measured by the Word Identification WRMT-III subtest?

RQ2b: What is the impact of Lexia on the ability to apply phonics rules to unfamiliar words of struggling readers in Year 2, as measured by the Word Attack WRMT-III subtest?

RQ2c: What is the impact of Lexia on the reading comprehension of struggling readers in Year 2, as measured by the Passage Comprehension WRMT-III subtest?

RQ2d: What is the impact of Lexia on the ability to read accurately and fluently of struggling readers in Year 2, as measured by the Oral Reading Fluency WRMT-III subtest?

This evaluation also includes four subgroup analyses, which answer the following RQs:

RQ3: What is the impact of Lexia on reading outcomes of struggling readers who are eligible for FSM, compared to those who are FSM eligible but receive usual reading instruction?

RQ4: What is the impact of Lexia on reading outcomes of struggling readers for whom English is an additional language, compared with EAL pupils who receive usual reading instruction?

RQ5: What is the impact of Lexia on reading outcomes of struggling readers who have SEND, compared with SEND pupils who receive usual reading instruction?

RQ6: How does Lexia differentially impact reading outcomes across struggling readers of different ability levels, given in quartile ranges?

The implementation and process evaluation (IPE) sought to answer the following questions in Table 3 based on the EEF's IPE guidance (EEF, 2022a).

Table 3: RQs with relevant IPE dimensions

IPE RQs	IPE dimensions
IPE RQ1: What are the expected activities, outputs, and outcomes of Lexia and how do these differ from business as usual?	Programme differentiation, monitoring of control / comparison groups (in a trial context)
IPE RQ2: In practice, which contexts and mechanisms mediate the intervention logic's causal pathways?	Moderators, mediators
IPE RQ3: To what extent was the intervention delivered as intended (including the selection of participating pupils, the administration of webinar-style training sessions, and the actual implementation of Lexia)?	Fidelity, dosage
IPE RQ4: What are the barriers and facilitators to the implementation of Lexia? Which barriers and facilitators, if any, were specifically or disproportionately experienced by FSM-eligible pupils? SEND pupils? EAL pupils?	Moderators, mediators
IPE RQ5: What changes or adjustments, if any, were made to the programme during its implementation? Why were these adjustments made?	Adaptation, fidelity

IPE RQs	IPE dimensions
IPE RQ6: How did settings divide tasks (i.e. facilitating Lexia sessions, weekly progress report monitoring, providing additional support to pupils) among their staff? What are the benefits and disadvantages of the different task divisions observed?	Quality
IPE RQ7: To what extent do school leadership, culture, and organisational structures influence the implementation of Lexia? What school-level policies and supports are in place to promote the use of digital technology in reading instruction and broader instructional practice?	Context, moderators
IPE RQ8: What does business as usual look like when it comes to reading instruction for struggling readers?	Monitoring of control groups (in a trial context)
IPE RQ9: How does Lexia differ from usual reading practice, which includes other targeted interventions that schools may implement as part of their business as usual?	Programme differentiation
IPE RQ10: To what extent did the mandatory training sessions and other available support enable practitioners to deliver the programme?	Moderators, mediators, quality
IPE RQ11: In what ways, if any, do practitioners use Lexia progress reports to make decisions about additional support provided to struggling readers?	Fidelity, quality
IPE RQ12: How do practitioners scaffold instruction in practice and to what extent does Lexia influence this process?	Fidelity, quality
IPE RQ13: To what extent did the programme improve the confidence and ability of practitioners to support struggling readers?	Participant responsiveness
IPE RQ14: To what extent does Lexia benefit disadvantaged pupils compared to business as usual?	Programme differentiation, quality
IPE RQ15: In what ways do the needs of disadvantaged readers differ from their non-disadvantaged peers? How does Lexia help address those needs?	Context, moderators
IPE RQ16: To what extent does Lexia result in positive or negative unintended consequences for settings, pupils, teachers, and other stakeholders?	Unintended consequences

Finally, there was a question on cost:

Cost RQ1: What is the cost of delivering Lexia and how does this compare to business as usual?

Ethics and trial registration

This evaluation was registered on the International Standard Randomised Controlled Trial Number (ISRCTN) Registry which is used to describe randomised controlled trials (RCTs) and efficacy trials at inception. The assigned registration number is: ISRCTN13603763.

The ethics and registration processes were in accordance with the ethics policies adopted by RAND Europe. The evaluation was approved by the RAND U.S. Human Subjects Protection Committee (HSPC).

The protocol was published on the EEF website (Merewood *et al.*, 2025); the Statistical Analysis Plan (SAP) was peer-reviewed and published on the EEF website (Speciani *et al.*, 2025).³

Schools were recruited to the trial by the project team. Interested schools were asked to read, sign, and return a copy of the Memorandum of Understanding (MoU), which explained what schools' participation in the trial would entail. See Appendix G for a copy of the MoU.

In addition, RAND Europe collected consent forms from school staff who participated in observations, interviews, or focus group discussions for the IPE. Furthermore, surveys with headteachers and practitioners included a privacy notice indicating to respondents that their participation was voluntary and that they could choose to withdraw at any time without penalty.

³ See: <https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/lexia-core5-reading-2022-2023-trial>

No member of the evaluation team had any conflict of interest with respect to the intervention or evaluation.

Data protection

Several teams were involved in controlling and processing data. RAND Europe acted as controllers during data collection, with Elklan (responsible for administering tests) acting as processors.

RAND Europe obtained personal data from settings as a data controller under the lawful basis of legitimate interest under the General Data Protection Regulation (GDPR). Legitimate interest is an appropriate basis because the data collected as part of this evaluation was used in ways that people would reasonably expect (i.e. for the benefit of improving the acquisition of reading skills in pupils) and that have minimal privacy impact. Legitimate interests apply where processing is necessary for the purpose of the legitimate interest pursued by the controller see GDPR Article 6 (1) (f) (GDPR, 2016) and for statistical and research purposes see GDPR Article 89 (GDPR, 2016). Special category data, such as information on pupils' SEND status, was processed for the purpose of scientific research and archiving as permitted under GDPR Article 9(2) (j) (GDPR, 2016). RAND Europe conducted a Data Privacy Impact Assessment (DPIA) that was reviewed and approved by RAND Europe's data protection officer prior to commencing and data collection in accordance with the Information Commissioner's Office (ICO) guidance on processing of child data. A Legitimate Interest Assessment (LIA) was also completed and signed off by RAND Europe's data protection officer.

The evaluation team archived data pertaining to the Lexia trial into the EEF's archive in the Office for National Statistics (ONS) Secure Research Service (SRS). All this data may potentially be linked to the NPD or to other data sources in the future for the purposes of research. When archived, this data has the potential to be used in further educational studies. No personal identifiers are submitted to the archive except for the Pupil Matching Reference (PMR), which is needed to link trial pupils with administrative data. The EEF archive is part of the ONS SRS, which significantly increases the security of the data. It can only be accessed by accredited researchers within the SRS. This fully complies with GDPR (2016) and the Data Protection Act 2018.

Personal and research data (e.g. names of pupils and school staff) was kept securely by the delivery team (LexiaUK and Queen's University of Belfast), the evaluation team (RAND Europe), and the testing sub-contractor (Elklan) for the duration of the study. This data, including test outcome data collected by Elklan, will be deleted by RAND Europe one year after the report has been published.

Parents/legal guardians, who acted as decisionmakers for individual pupils, were allowed to request that their child's data be removed from the dataset (if data was already collected) or to not participate in the evaluation (if data had not been collected). In the interest of informed consent, parents/legal guardians were provided with information sheets and withdrawal forms (see Appendix H). To withdraw a child from the trial, parents/legal guardians were able to return the withdrawal form provided or email the evaluation team directly. Parents/legal guardians were able to request this at any point during the intervention and evaluation. If a child's participation was withdrawn, the evaluation team did not collect data from the child or deleted any data previously collected, as needed.

In addition, RAND Europe collected consent forms from school staff who participated in observations or interviews for the IPE. Furthermore, surveys with headteachers, teachers, and TAs included a privacy notice indicating to respondents that their participation is voluntary and that they can choose to withdraw at any time without penalty (see Appendix H).

No member of the evaluation team has any conflict of interest with respect to the intervention or evaluation.

Project team

Delivery team

- Maria Cockerill, Project Lead, Queen's University of Belfast.
- Joanne O'Keefe, Project Manager, Queen's University of Belfast.
- Rob Kay, Director, LexiaUK.
- Norman Robinson, Chief Operating Officer, LexiaUK.

Evaluation team

- Elena Rosa Speciani, Project Lead, RAND Europe.
- Rebecca Mead, Project Manager, RAND Europe.
- Miguel Subosa, Project Manager, formerly RAND Europe.

Fieldwork and analysis team

- James Merewood, RAND Europe.
- Louis Hodge, RAND Europe.
- Emily Hutton, RAND Europe.
- Anna Brian, RAND Europe.
- Katie Sykes, RAND Europe.
- Lili Xu, RAND Europe.
- Ben Senator, formerly RAND Europe.
- Mark Reed, formerly RAND Europe.
- Anirudh Agarwal, formerly RAND Europe.
- Emily MacLeod, formerly RAND Europe.
- Fin Oades, formerly RAND Europe.
- Giovanni Amodeo, formerly RAND Europe.
- Sarah Angell, formerly RAND Europe.
- Bhavya Singh, formerly RAND Europe.

Testing partner

- Elklan

Methods

Trial design

Table 4: Trial design

Trial design, including number of arms		Two-arm, cluster RCT
Unit of randomisation		School
Stratification variable (s) (if applicable)		School in an EIA versus school not in an EIA
Primary outcome	Variable	Reading proficiency
	Measure (instrument, scale, source)	WRMT-III score: Composite of three subtests used as outcome (Word Identification, Word Attack, Passage Comprehension)
		Scale is given as a standardised score, with an average range (25th–75th percentile) of 270–330
Secondary outcome(s)	Variable(s)	Skill-specific reading outcomes
	Measure(s) (instrument, scale, source)	WRMT-III age-standardised scores for each of the following three subtests: Word Identification, Word Attack, Passage Comprehension, and Oral Reading Fluency
	Variable	Reading proficiency
Baseline for primary outcome	Measure (instrument, scale, source)	WRMT-III score: Composite of three subtests used as outcome (Word Identification, Word Attack, and Passage Comprehension)
		Scale is given as a standardised score, with an average range (25th–75th percentile) of 270–330
Baseline for secondary outcome(s)	Variable	Skill-specific reading outcomes
	Measure (instrument, scale, source)	WRMT-III age-standardised scores for each of the following three subtests: Word Identification, Word Attack, and Passage Comprehension
		In the case of Oral Reading Fluency, WRMT-III composite is used as the baseline measure

This evaluation was designed as a two-armed, cluster RCT. Randomisation took place at the school level, with EIA status used as a stratification variable. This means that an equal number of schools from EIAs were included in the treatment and control conditions. Since the programme is funded by the Accelerator Fund, which targets schools in disadvantaged locales, 50% of recruited schools were intended to be in EIAs. This target was reached, with 53.7% of the schools recruited coming from EIAs.

Schools were assigned to either treatment or control. The control condition represented business as usual while the treatment condition represented schools receiving access to the Lexia programme. Before randomisation, schools identified pupils to receive the treatment. Teachers selected struggling readers in Year 1 (the year before implementation) using selection criteria developed by the delivery team. While further details on school and pupil selection can be found in the ‘Participant selection’ section, the trial sought to mimic usual practice by having teachers select participating pupils based on their own judgement with guidance provided by LexiaUK. Therefore, for the purposes of this trial, pupils were classified as ‘struggling readers’ if teachers felt that the pupils fell below age-normative standards for reading proficiency.

A split-cohort design was adopted due to constraints in programme delivery and school recruitment capacity: it would have been challenging for the delivery team to recruit and deliver the programme to the number of schools necessary for a well-powered trial if the programme had been administered with a single cohort. In total, we collected data from 223 schools over two cohorts: 112 assigned to treatment, and 111 assigned to control. Cohort 1 contained 112 schools, with half

assigned to treatment and half assigned to control (i.e. 56 treatment schools and 56 control schools in each cohort). Cohort 2 contained 111 schools, with 56 assigned to treatment and 55 to control.

The primary outcome was pupils' reading proficiency outcomes, measured using pupils' scores in the WRMT-III. The composite reading score was comprised of the following WMRT-III subtests: Word Identification; Word Attack; and Passage Comprehension. The use of the composite reading score for WMRT-III mirrors the approach used in the efficacy trial (Tracey *et al.*, 2022) and is recommended by the developer (along with Word Comprehension) as a broad measure of global reading achievement. The Word Comprehension subtest was also considered but was excluded based on the efficacy trial reporting that it tested similar capabilities already captured by the Passage Comprehension subtest. WRMT-III composite scores were used for both baseline and endline testing (Tracey *et al.*, 2022).

In the SAP (Speciani *et al.*, 2025), we stated that the outcome variable would be the composite of four subtests: the three listed above; and Oral Reading Fluency. However, during baseline data collection in the first cohort, test administrators from Elklan reported that the Oral Reading Fluency subtest was at times too difficult for pupils and risked distressing some of them. The EEF, LexiaUK, Queen's University of Belfast, and the evaluation team made the decision to stop collecting it at baseline. Testers administered this subtest at endline, but some pupils were still unable to complete the test. Given concerns that bias would be introduced by having a primary analysis comprised only of pupils able to complete the Oral Reading Fluency subtest, the evaluation team chose to drop this problematic subtest from the primary analysis and use a composite of the other three subtests (Word Identification, Word Attack, and Passage Comprehension) at baseline and endline. Data was still collected for pupils able to complete the test, enabling us to perform a sensitivity analysis (see 'Statistical analysis' section below) using all four subtests, providing results more comparable to the efficacy trial, which uses a composite of all four subtests at endline (Tracey *et al.*, 2022).

The secondary outcome was pupils' scores in the following WRMT-III subtests: i) Word Identification; ii) Word Attack; iii) Passage Comprehension; and iv) Oral Reading Fluency. Further details on the measures can be found in the 'Outcome measures' section below.

Participant selection

School selection

The trial was implemented across 224 schools, recruited nationwide across England. Since the intervention is funded through the Accelerator Fund, recruitment was targeted with the goal that at least 50% of these schools would be located in EIAs—educationally underperforming areas that have been prioritised by the British Government for further support and funding.⁴ At randomisation, 53.7% of schools were in EIAs.

To be considered for participation in the trial, schools needed to respond to a publicly available request for Expressions of Interest (EOIs). This request for EOIs was published on dedicated pages on the respective websites of the EEF and LexiaUK, the programme provider. The Queen's University of Belfast led school recruitment for this trial. The Queen's University of Belfast team contacted schools with the school information sheet and MoU documents and followed up on EOIs received via the EEF and LexiaUK websites.

Schools that expressed an interest were screened against the following eligibility criteria:

- had a minimum of eight pupils (and a maximum of 16 pupils) that would be eligible for Lexia;
- not currently involved in any other trial by the EEF focusing on Year 2;

⁴ The UK Government's levelling up white paper (HM Government, 2022) identified 55 EIAs. EIAs correspond to local authorities that showed sustained low educational outcomes across Key Stage 2 and Key Stage 4 from 2017 to 2019.

- had not held a Lexia licence in the 12 months prior to July 2022 (for Cohort 1) and July 2023 (Cohort 2); and
- able to provide the necessary IT equipment required to support the intervention.

Once deemed eligible, and after completing the baseline assessment, schools were randomised into the control or treatment groups in accordance with the process described in the 'Randomisation' section.

School recruitment strategy

Schools that expressed an interest in participating in the trial were contacted by Queen's University of Belfast, who also screened them against the previously indicated eligibility criteria. Recruitment was carried out nationwide to maximise geographical spread in the trial sample. Furthermore, schools in EIAs were prioritised for recruitment to ensure that at least 50% of participating schools come from these areas. This was in keeping with the requirements of the DfE's Accelerator Fund, which financed this evaluation.

Following the screening process, Queen's University of Belfast distributed and collected MoUs from eligible schools. Afterwards, the evaluation team (RAND Europe) provided schools with information sheets for parents/legal guardians, a privacy notice, a document outlining guidance for pupil selection (from LexiaUK), and a data collection template in which schools inputted requested data on pupils nominated to participate in the trial.

Pupil selection

For the purposes of the evaluation, Lexia was administered to Year 2 pupils in schools across different regions in England. While Lexia can be delivered to pupils of any age group, the evaluation team selected Year 2 pupils as the population to align with the efficacy trial, which used this age group for its sample. This facilitates comparison between the findings of the efficacy and effectiveness trials. Year 2 pupils were identified as the population for the EEF initial efficacy trial to test the assumption that struggling readers in Year 2 would benefit most from Lexia, following their standard Year 1 phonics check and completion of the Year 1 literacy curriculum.

The trial participants were struggling readers. Eligible pupils were identified by their classroom teacher when they were in Year 1, in the term before they went into Year 2, using written guidance provided by LexiaUK on the selection of struggling readers. In each school, a minimum of eight and a maximum of 16 struggling readers received Lexia. The written pupil selection guidance laid out the following criteria:

- Pupils are able to stay on task at a computer for 20–30-minute sessions.
- They are available to attend four to five Lexia sessions every week during the school's planned delivery schedule.
- Hearing-impaired pupils may struggle if their amplification device does not enable them to hear spoken language and isolated sounds. Therefore, deaf pupils may not be suitable candidates for Lexia use.
- Visually impaired pupils might benefit from Lexia if they can read print on a computer screen or tablet. However, blind pupils may not be suitable candidates for Lexia use.
- For pupils with physical disabilities such as paralysis or cerebral palsy, specialist devices provided by the school and set-up by assistive technology specialists may help access to Lexia.

These struggling readers were identified by teachers, based on existing evidence that teachers are able to successfully identify appropriate pupils when armed with the correct guidance (Hoge and Coladarci, 1989). Each participating school nominated one Year 2 teacher and one TA, who were given written pupil selection guidance by LexiaUK and asked to share with the Year 1 teacher for selection of pupils (when the pupils were in Year 1).

The three-year Lexia licence package included up to 50 pupil licences. In this trial, as per business as usual practice, (treatment) schools had administrative access from the beginning of their licence package. They had the functionality to

freely add new pupils, staff, and classes from the outset, although LexiaUK guidance and training advised them to start with a manageable priority group (in this case, their Year 2 trial pupils). After all training was delivered and fidelity of implementation was achieved, schools were in a position to scale to more pupils over time. An implementation plan was developed over the course of the rollout period.

To help support fidelity, LexiaUK created ‘the EEF Study Class’ grouping of the selected trial pupils so that these are already set-up for schools on their dashboards. This reminded schools that they should not redistribute licences for the pupils in the study over the course of the trial period.

Outcome measures

Baseline and endline primary measures

The baseline measure was WRMT-III (Woodcock, 2011), which measures the reading ability of people between the ages of 4.6 to 79.11. The same version of the WRMT-III is used for both baseline and endline. This is a commercial instrument and is not included in the report. Tests were administered by Ekklan who were blind to the randomisation status of schools and were independent from the evaluation team.

The WRMT-III includes the following subtests: Word Identification; Word Attack; Passage Comprehension; Word Comprehension; and Oral Reading Fluency. In Word Identification, pupils are required to read a list of words which increase in difficulty. Word Attack asks the pupil to read a list of nonsense words of increasing difficulty. Passage Comprehension measures the ability of pupils to identify missing words in a passage of text, which in turn requires them to use comprehension skills to understand the context of the written passage. Word Comprehension measures the ability of pupils to cognitively process words through antonyms, synonyms, and analogies. Oral Reading Fluency assesses the ability of pupils to read aloud accurately and fluently. These subtests capture different elements of reading proficiency:

- **Word Identification subtest.** Tests the ability to read and pronounce words.
- **Word Attack subtest.** Tests the ability to apply phonics.
- **Passage Comprehension subtest.** Tests reading comprehension.
- **Oral Reading Fluency subtest.** Captures the ability to read accurately and fluently.

The Word Comprehension subtest was available for consideration but was excluded based on the efficacy trial reporting that it tested similar capabilities to those already captured by the Passage Comprehension subtest. Therefore, including the Word Comprehension subtest was felt to provide little additional information. Unlike what was laid out in the SAP (Speciani *et al.*, 2025), due to difficulties in conducting the Oral Reading Fluency test (see the ‘Trial design’ subsection above), this was also excluded from the composite measure but retained as a secondary outcome.

WRMT-III was administered at baseline and endline. This allowed us to directly compare endline scores of pupils assigned to control and treatment schools while controlling for their baseline scores. Test scores from the individual WRMT-III subtests are standardised with an average range (i.e. the 25th–75th percentile range) of 90–110. Additional information provided as part of the test includes percentile ranks, confidence intervals (CIs), age equivalents, and relative performance indices. The test scores provided by WRMT-III are useful for evaluating struggling readers, screening for reading readiness, and evaluating programme effectiveness, which is why it was selected for the previous Lexia trial (Tracey *et al.*, 2022). The final composite test score—that is, the sum of the age-standardised scores on each of the three subtests—was our primary outcome measure, resulting in a possible score range of 176 to 435, as stated in the efficacy trial (see Tracey *et al.*, 2022).

The WRMT-III was selected as the main outcome measure for two reasons: i) psychometric soundness; and ii) comparability with the efficacy trial. According to Pearson, the WRMT-III (which was normed against individuals aged 4.6 through 79.11) registered a high level of internal consistency for its age-standardised Total Reading Cluster scale ($\alpha = 0.98$) and a test-retest

reliability ranging from $r^2 = 0.51$ to 0.95 .⁵ Furthermore, it has demonstrated correlations with tests of similar outcomes, such as: the Kaufman Test of Educational Achievement, Second Edition (KTEA™-II); the Wechsler Individual Achievement Test, Third Edition (WIAT-III); and the Woodcock-Johnson III Tests of Achievement (WJ-III ACH) (Woodcock, 2011). The WRMT-III has also been shown to capture developmental change in reading outcomes across age groups. These correlations with similar scales and the test's ability to capture developmental change attest to the WRMT-III's construct validity. In addition to these psychometric properties, the WRMT-III was also used as the main outcome metric in the efficacy trial; using it as the main outcome metric for the effectiveness trial as well will facilitate comparison between the findings of the former and the latter. This measure of reading includes skills relevant to the skills included in the Lexia theory of change (Figure 1) such as vocabulary, comprehension, and fluency, meaning that it captures the targeted outcomes of the intervention.

Secondary outcomes

Our secondary outcomes came from the individual, age-standardised subtest scores on the WRMT-III. The subtests chosen were intended to test basic skills (Word Identification and Word Attack), reading comprehension (Passage Comprehension), and reading fluency (Oral Reading Fluency). Using these subtests in secondary outcomes allowed us to test if, and the extent to which, Lexia was more effective at improving specific reading skill domains over others, or whether the effect of Lexia is homogeneous across the WRMT-III subtests. As previously stated, we omitted the Word Comprehension subtest, following the approach of the efficacy trial, which found the skills captured by this subtest to repeat most of the same skills captured by the Passage Comprehension subtest.

The baseline measures in the secondary measure analyses correspond to the measures used at endline, with the exception of the Oral Reading Fluency subtest. As this subtest was not conducted at baseline, we use the total WRMT-III score as the baseline measure in this analysis.

The subtest scores are given in an age-standardised form, thus already accounting for differences in pupils' ages when they take the test.

Sample size

The minimum detectable effect size (MDES) for this study was calculated using a two-level random assignment design. In calculating MDES, a number of assumptions were used: two-tailed; with 50:50 allocation; alpha of 0.05; and power at 0.8. All MDES calculations were made using PowerUp! (Dong and Maynard, 2013).

Our assumptions at the protocol stage were based on the efficacy trial results ($g=0.08$) and the mean cluster size of 12 used in the efficacy trial (Tracey *et al.*, 2022). We also considered the range of potential effect sizes reported in two meta-analyses of computer-assisted reading instruction programmes, which reported aggregate effect sizes of 0.11 and 0.1316 (Cheung and Slavin, 2012; Soe *et al.*, 2000). These aggregate effect sizes were relatively close to the computed overall effect size in the efficacy trial. Based on these calculations, we proposed a sample of 224 schools (combined across the two cohorts) with a power to detect an effect size of 0.115. After randomisation (i.e. after Cohort 1 and Cohort 2 were randomised) we had 222 schools in the sample (two schools fewer than expected) with an average cohort size of 12 pupils per school. Keeping the assumptions the same from the protocol stage, changing only the number of schools, at randomisation stage we have the power to detect an MDES of 0.116.

The efficacy trial detected differential impact on FSM-eligible pupils, compared to non-FSM-eligible pupils; for FSM-eligible pupils, the calculated effect size was slightly larger at 0.18 (compared to 0.08 for all pupils). However, since the efficacy trial was not powered to detect FSM subgroup effects, these findings cannot be considered conclusive. Therefore, we aimed to power this trial to isolate treatment effect on FSM-eligible pupils. FSM-specific calculations at recruitment and randomisation are presented in Table 8, using statistical coefficients for FSM-eligible pupils reported in the efficacy trial. At this sample size, the MDES for the FSM-eligible cohort of pupils was 0.159 at protocol stage and 0.160 at randomisation.

⁵ See: www.pearsonclinical.co.uk/en-gb/Store/Professional-Assessments/Academic-Learning/Woodcock-Reading-Mastery-Tests%2C-Third-Edition/p/P100009264

Randomisation

Randomisation was conducted with a 50:50 allocation to treatment and control, stratified by school's EIA status (i.e. EIA vs non-EIA). Randomisation was conducted in Stata (StataCorp LLC, College Station, Texas, USA) by a member of the evaluation team who was blind to treatment allocation using the *randtreat* package, with misfits independently randomly allocated across strata (Carril, 2017). The code used to assign schools to treatment or control was communicated to the implementation team in a PDF to prevent editing and was recorded in a 'Stata do' file.

Randomisation was conducted by RAND Europe immediately before the school summer holiday period (i.e. last week of July) for each cohort. Randomisation was conducted by a researcher outside the research team that was blind to allocation.⁶ Before randomisation took place, schools identified target pupils and arranged test dates (from June to July) with Elklan. Once schools allowed identified pupils to take part in baseline testing and returned pupil data, RAND Europe considered them eligible for the evaluation. At a pre-specified date, all schools eligible for randomisation were randomised by RAND, who communicated with all schools the details of allocation. RAND Europe also communicated next steps to intervention schools.

Statistical analysis

Primary analysis

Our primary analysis followed the intention-to-treat (ITT) principle, as outlined in the EEF's statistical analysis guidance (EEF, 2022b). We used the composite WRMT-III score as our dependent variable, and a binary variable indicating whether the observation was included in the treatment or control as an independent variable. Under an ITT approach, analysis includes all randomised settings and baselined struggling readers, grouped according to random assignment, regardless of programme compliance or treatment dosage. It is an inherently conservative approach, estimating the average effect of offering the intervention, and is key to ensuring an unbiased analysis of intervention effects in line with the EEF's statistical analysis guidance (EEF, 2022b).

We use a multilevel modelling approach. The first level being the unit of analysis (the pupil) and the second level the unit of randomisation (the school). This approach reflects the trial design and nested nature of the data, in line with the EEF statistical analysis guidance (EEF, 2022b). It appropriately clusters the error term at the unit of randomisation to ensure appropriate and unbiased CIs are estimated.

Specifically, we estimate a two-level random-intercept model, outlined below in Equation 1. This is known as a random-intercept model because the school-specific intercepts for each school j , $(\beta_{0j} = \beta_0 + u_j)$ vary randomly, where the school-level residuals (u_j) are assumed to be normally distributed ($u_j \sim N(0, \sigma_u^2)$). Thus, the intercepts for each school are distributed as $\beta_{0j} \sim N(\beta_0, \sigma_u^2)$. The model controls for pre-test (baseline) attainment, as measured by baseline WRMT-III scores, and estimated fixed effects for the stratification variable (a binary indicator for if a school is in an EIA) at the school level.

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

where:

- Y_{ij} is the WMRT-III composite reading score for child i in setting j , at endline;
- β_0 is the cluster-level intercept for the slope of a predictor on reading proficiency;

⁶ Meaningless identifiers were used in the place of school names. The researcher randomised the meaningless identifiers and shared the resulting allocation to the team who had access to the key to re-identify the schools.

- Lexia_j is a binary indicator of the setting assignment to intervention [1] or control [0];
- Z_j is a vector of school-level characteristics, i.e. the stratifying variable of whether a school is in an EIA (as used for randomisation), and cohort;
- X_{ij} is a vector of child-level characteristics for child i in setting j , or more, specifically the baseline WMRT-III composite reading score and a factor variable indicating whether the child was part of Cohort 1 or Cohort 2,
- u_j are the school-level residuals; and
- e_{ij} are the pupil-level residuals.

The coefficient τ was the outcome of interest, as an estimate of the conditional effect of treatment on endline WMRT-III composite reading score. We subsequently calculate a standardised effect size using Hedges' g for τ .

All analyses were conducted in R (R Core Team, 2025). Multilevel models were fitted using the *lme4* package (Bates *et al.*, 2015).

Multiple cohort analysis

We consider two approaches to analysing the data from the two different cohorts of Year 2 pupils. In our primary analysis model, we combine the data across cohorts into a single sample and include a factor variable for cohort. This approach retains the full analytical sample, increasing statistical power and reducing the risk of type II errors. However, this approach assumes that the observable and unobservable characteristics of the pupils, and delivery of Lexia, did not differ across the two cohorts. Since we randomised at the school level *within* an academic year's cohort but not *between* the cohorts, any cross-cohort differences could confound the estimated effect of Lexia. The inclusion of a child-level indicator for cohort partially controls for such differences.

In a sensitivity analysis, we treat each cohort as a separate trial to maintain the independence of the two datasets and avoid assumptions about inter-cohort similarity. While this reduces the risk of confounding but provides less statistical power than the first approach due to a decreased analytical sample size. This is a particular concern for subgroup analyses in which sample sizes are by nature smaller. Here, we conduct a meta-analysis to estimate an aggregate weighted effect size of Lexia across both cohorts. This enables us to both reflect on the inter-cohort differences in Lexia's effects as well as provide a sensitivity check to the overall estimates.

Secondary analysis

Our secondary analysis follows the considerations made in the primary analysis (above) and the ITT principle, using the subtest scores of WRMT-III as the outcome of interest.

We used the same multilevel modelling approach as in the primary analysis. That is, more specifically, we estimated a two-level random-intercept model (see 'Primary analysis' section above for justification and formal model specification), where the dependent outcome variable differs in each regression. The first regression focused on the impact of Word Identification, the second on Word Attack, and the third on Passage Comprehension, and fourth on Oral Reading Fluency. In each regression, the independent variables matched those in the primary analysis: a treatment indicator; EIA status; and cohort and respective subtest scores at baseline.

Given the multiple secondary outcomes, we applied a Romano-Wolf correction to correct for potential over-rejection of null hypotheses arising from multiple hypothesis testing. As the WMRT-III is a composite measure of reading proficiency, its subtests are likely to be moderately correlated. In such circumstances, the Bonferroni correction can be overly conservative, so the Romano-Wolf correction provided a more appropriate method. This correction accounts for the dependent nature of the test statistics and provided a strong control against the family-wise error rate (Clarke *et al.*, 2019), as recommended in the EEF's statistical analysis guidance (EEF, 2022b).

All analyses were done in R. Multiple hypothesis testing corrections were made using *crctStepdown* (Watson, 2024).

Analysis in the presence of non-compliance

The primary analysis captured the average effect of offering the Lexia intervention on an ITT basis. However, we also examined treatment effects in the presence of non-compliance. To do this, we include analysis on the impact of compliance on our primary outcome variable.

In consultation with the EEF and the Lexia delivery team, we defined a compliance metric and a dosage metric.

Compliance

The minimum compliance was defined as completing at least 60 minutes a week for a minimum of 12 non-consecutive weeks equal to completing a minimum of 720 minutes over a 24-week period. The expectation was that schools likely used the programme much more than this: guidance to teachers suggests that four 30-minute sessions take place each week (120 mins), for struggling readers. As such, the compliance metric was a binary variable indicating if a pupil used (or did not use) Lexia for a minimum of 720 minutes. This data was computed by the delivery team using usage reports generated from the software.

We calculated the Complier Average Causal Effect (CACE) through a two-stage least squares (2SLS) instrumental variable approach. The first stage of this approach modelled the impact on the binary compliance indicator of treatment allocation with the same pupil- and school-level covariates and multilevel structure using a linear probability model. This provided an estimate of how assignment of pupils and settings to receive Lexia encouraged uptake of the intervention.

Results from this first stage effectively provided an overall ‘compliance rate’ and was estimated using the following equation:

$$C_{ij} = \beta_0 + \text{Lexia}_j\tau + Z_j\beta_1 + X_{ij}\beta_2 + u_j + e_{ij} \quad (\text{Equation 2})$$

where:

- C_{ij} is the binary compliance indicator for pupil i in setting j ;
- β_0 is the cluster-level intercept;
- Lexia_j is a binary indicator of the setting assignment to intervention [1] or control [0];
- X_{ij} is a vector of pupil-level characteristics;
- Z_j is a vector of school-level characteristics;
- u_j are the school-level residuals; and
- e_{ij} are the pupil-level residuals.

The second stage of this approach follows Equation 1, but with the primary outcome variable being regressed on the predicted compliance from the first stage ($\widehat{C}_{ij}$) in place of the treatment assignment (Angrist and Krueger, 1991; Angrist, 2006). The Hedges’ g derived from this model was similarly calculated to provide an estimate of the CACE. The results of each of these two models allowed us to discern the degree to which compliance with the Lexia intervention improved outcomes for pupils. This analysis was conducted for the primary outcome only.

Given the relatively high number of minutes spent on Lexia recorded in the trial, the initial definition of compliance included a very large proportion of the analysed treated group (97.4%). We also noted that an overwhelming proportion of pupils completed more minutes than what was defined as compliant.

Dosage

As part of the SAP (Speciani *et al.*, 2025), we also wanted to explore whether higher thresholds of compliance led to higher estimates of the treatment effect. We therefore conducted an additional dosage analysis, mirroring the 2SLS method for the binary compliance measure but instead using a continuous outcome measure of Lexia sessions (as measured by time logged on the Lexia platform) in the first stage.

To account for potential non-linearities we also looked at the incremental benefits of spending more time on Lexia—the impact of receiving 2x, 3x, 4x, and 5x the recommended Lexia dosage. This analysis followed the structure of the compliance analysis described above, increasing the threshold for the binary compliance measure.

In addition, to understand the extent to which schools may have differed in the way they had distributed Lexia sessions on a weekly basis, we reported descriptive statistics (e.g. range, mean) on overall use.

Missing data analysis

Attrition across both trial arms was explored to identify any potential bias which exists due to attrition based on specific characteristics of schools or pupils. We calculate the attrition rate by observing how many of the pupils present at baseline assessment were still present at endline assessment.

Missingness for the primary outcome accounts for greater than 5% of observations, so we conducted a systematic analysis of missingness as a function of observable characteristics contained within the original dataset and the NPD (for variables such as FSM eligibility). To establish whether systematic differences exist between those who are ‘missing’ and those who are not, we modelled missingness at follow-up (defined as pupils with missing primary outcome data at endline) as a function of the baseline covariates included in the crosstabulations. This took the form of a logistic regression model, using a binary outcome variable denoting missingness at endline (where 1=missing; 0=complete), and mirrored the multilevel structure of the models used in the main analysis, with pupils nested within settings.

Based on the outcome of this analysis, we utilised a multiple imputation approach, generating replacement values for observations with missing primary outcome data based on characteristics available at baseline. We used a fully conditional specification (FCS) method, and averages of 20 imputations were used to calculate revised estimates to reduce standard errors and improve statistical power. This method of imputation is seen to be consistent and efficient in clinical trials (Jakobsen *et al.*, 2017), as it accurately estimates values for missing data. We used multilevel Multivariate Imputation by Chained Equations (MICE). This is done by first generating several datasets, each containing a plausible value for the missing data; next, the primary analysis is conducted on each of the datasets which have been created individually (Sterne *et al.*, 2009). Finally, the results from each of the analyses conducted (on each of the filled datasets) are pooled into one multiple imputation result which provides insight into the true effect of the trial, even in the presence of data missing not at random.

Subgroup analyses

FSM-eligible pupils

As discussed in the ‘Sample size’ section above, our trial was powered to identify an effect on the subgroup of FSM pupils. This subgroup analysis formed an important part of the study (specifically, RQ3) since the efficacy trial found evidence of an effect on FSM pupils but was not highly powered enough to suggest a causal link. The efficacy trial found two months’ additional reading progress for FSM pupils, compared to their non-FSM counterparts, for whom one month’s additional reading progress was observed.

This subgroup analysis utilised a binary variable indicating whether a pupil is eligible for FSM. To identify whether a pupil is FSM eligible, we used the variable *EVERFSM_6_P* from the NPD which required us to link our trial data with the NPD in the ONS SRS. To provide evidence of the impact of Lexia on FSM-eligible pupils, we estimate multilevel models consistent with Equation 1 with the addition of the FSM indicator in the vector of pupil characteristics (X_{ij}) and an interaction term between the treatment and the FSM indicator.

EAL pupils

In line with RQ4, the second subgroup analysis undertaken in this project focused on EAL pupils. This subgroup analysis provided further understanding on whether Lexia is appropriate for pupils who need a high level of additional scaffolding (i.e. supplementary learning activities outside Lexia) from their teachers. Scaffolding is a technique whereby teachers provide temporary support to a pupil depending on their level of mastery of a subject or skill, and then slowly withdraw the support over time (Van der Stuyf, 2002). In reading provision, this may take the form of segmenting sections of text and considering each in turn or first reading through the text before returning to comprehension tasks. This type of teaching can help to transfer foundational knowledge from the teacher to the pupil in a more structured way (Ge and Land, 2003; Van der Stuyf, 2002; Vygotsky, 1978).

To understand the impact of Lexia on EAL pupils, we used the *FirstLanguage* variable in the NPD. This is a categorical variable with six categories that we condense into three, denoting whether a child's language in the home and community is English, not English, or unknown. We again ran multilevel models with the inclusion of the categorical variable indicating whether the pupil is EAL or believed to be EAL, along with an interaction term between the EAL variable and the treatment variable. This is informative about the additional impact of being treated as an EAL pupil.

SEND pupils

To address RQ5, the third subgroup analysis undertaken focused on SEND pupils. In acknowledgement that SEND pupils have been identified in other research on teaching and literacy (Tohara *et al.*, 2021), this subgroup analysis aimed to establish whether Lexia is appropriate for SEND pupils, by comparing the effect of Lexia on SEND pupils who received it compared to SEND pupils who did not receive the intervention.

To understand the impact of Lexia on SEND pupils, we used the *SENprovision* variable in the NPD. This is a categorical variable with three categories (No SEND; SEND support; Education, Health, and Care Plan [EHCP]) which we condensed into two, denoting whether a child is considered SEND or not SEND. We included both the binary SEND variable and an interaction term between the binary SEND variable and the treatment indicator variable in the model, allowing us to establish the differential impact of being treated for SEND pupils.

Lower levels of initial reading attainment

Our final subgroup analysis examined the ability of pupils using their scores in the WRMT-III baseline outcome variable (discussed above, specifically to answer RQ6). This subgroup analysis determined whether Lexia was more successful at providing help to pupils within a certain range of reading proficiency levels. To make comparisons between quartiles meaningful, we examined the effect of Lexia on a range of increasing thresholds of ability, as measured by the WRMT-III baseline test scores. First, we analysed the impact of Lexia on a subsample restricted to pupils who scored below a 10% threshold (i.e. have scored in the bottom 10% of the maximum possible test score). We then increased this threshold by 10% at a time, including pupils who are below a 20%, 30%, 40%, and 50% threshold, respectively. The number of pupils in the bottom 50%, 558 is slightly higher than the number in the top 50%, 511. This allowed us to identify the pupil ability threshold at which the Lexia treatment became less effective; though it is important to note that the sample size is inherently considerably smaller for this analysis. For robustness, we generated the reverse scale with a series of regressions restricted to a subsample of >90%, >80%, >70%, >60%, and >50%. This captured the full sample within the ability subgroup analysis and could help to inform future decisions on the selection of pupils who receive the Lexia package.

Additional analyses and robustness checks

Cohort analysis

As discussed above in the ‘Multiple cohort analysis’ subsection above, we included a robustness check to look at the impact of an alternative approaches to calculating an aggregate estimate for both cohorts has on the final impact estimate. We treated each as a separate trial, using random effects meta-analysis to combine results allowing greater variability between cohorts than the primary approach.

Oral fluency

Given that we made the decision to exclude the Oral Reading Fluency test from the primary analysis due to difficulty collecting the data at baseline (see ‘Outcome measures’ section), the composite of the four subtests (Word Identification, Word Attack, Passage Comprehension, and Oral Reading Fluency) used at endline in the efficacy trial (Tracey *et al.*, 2022) is not available at baseline, and using it as the endline outcome measure risked introducing bias. We therefore replaced the model using the composite of the four subtests at baseline and endline proposed as the primary analysis in the SAP (Speciani *et al.*, 2025) with one using three subtests at baseline and endline. However, the data collected from this test at endline enabled us to conduct a sensitivity analysis repeating the primary model using the composite of all four subtests as the endline outcome variable, keeping the composite of three subtests excluding Oral Reading Fluency as the baseline outcome measure.

Estimation of effect sizes

We calculated effect sizes for cluster-randomised trials as outlined in the EEF evaluator guidance (EEF, 2022b), and adapted from (Hedges, 2007):

$$ES = \frac{(\bar{Y}_T - \bar{Y}_C)_{adjusted}}{\sqrt{\frac{(n_1 - 1)sd_1^2 + (n_2 - 1)sd_2^2}{n_1 + n_2 - 2}}}$$

Where $(\bar{Y}_T - \bar{Y}_C)_{adjusted}$ is the mean difference between the intervention and control group adjusted for baseline characteristics and $\sqrt{\frac{(n_1 - 1)sd_1^2 + (n_2 - 1)sd_2^2}{n_1 + n_2 - 2}}$ is an estimate of the pooled unconditional population standard deviation (SD).

The pooled unconditional SD is the weighted average of SDs of treatment and control (Coe, 2002). The pooled unconditional SD across the two trial arms was used in the denominator, as we assumed the SDs of both the treatment and control groups are drawn from the same underlying population distribution.

From the primary outcome model, we took each group’s adjusted mean and variance to calculate the effect size. This variance was the total variance (across both pupil and school levels, without any covariates, as emerging from a ‘null’ or ‘empty’ multilevel model with no predictors). The *ES* therefore represents the proportion of the population SD attributable to the intervention (Hutchinson and Styles, 2010). A 95% CI for the *ES*, which takes into account the clustering of pupils in schools, was reported. Effect sizes were calculated for each of the models estimated.

Estimation of intracluster correlation coefficient (ICC)

The ICC is a crucial metric for trials involving clusters. It quantifies the fraction of variance in a specific outcome attributable to differences between clusters (such as settings), rather than variance occurring within these clusters.

The ICC applied in the power calculations is detailed in the ‘Sample size’ subsection above. At the protocol stage (refer to the protocol for further details; Merewood *et al.*, 2025), this is set at 0.06 for all pupils and 0.075 for FSM-eligible pupils. This value is derived from the ICC observed in the EEF previous efficacy trial.

The ICC at the analysis stage focused on the primary outcome measure. Its calculation involves two approaches: i) using the model corresponding to Equation 1; and ii) employing a model akin to that in Equation 1 (clustering pupils in schools) but without any covariates, i.e. only containing the outcome and treatment indicator. The ICCs at analysis stage are presented in Table 5.

Table 5: Calculated ICC at analysis stage

Outcome	All pupils (95% CI)	FSM-eligible pupils (95% CI)
ICC (primary model)	0.071 (0.039, 0.101)	0.063 (0.005, 0.120)
ICC (empty model)	0.165 (0.126, 0.197)	0.130 (0.063, 0.208)

ICCs at the analysis stage were estimated using the R package *performance* (Lüdtke *et al.*, 2021).

IPE

Research methods

A mixed-methods approach was used to answer the RQs listed in the ‘Evaluation objectives’ subsection in the ‘Introduction’ section above, allowing for triangulation, and detailed comparison of treatment and control schools. Data collection methods for the IPE included baseline and endline surveys, key informant interviews, focus group discussions, and observations. These methods helped develop a fuller understanding of causal pathways that affected programme impact. In other words, the IPE complemented the impact evaluation: while the latter estimated the magnitude of the intervention’s effect, the IPE explained how that effect came about.

Table 6 below maps each data collection method against the RQs that it sought to answer, as well as the relevant IPE dimensions.

Table 6: IPE methods overview

Research methods	Data collection methods	Mode	Participants	Data analysis methods	IPE RQs addressed	Relevant IPE dimension
Surveys	Questionnaires: baseline survey	Online	Headteachers and nominated Year 2 teachers from participating schools	Descriptive statistics Thematic analysis	1, 7, 8	Programme differentiation, monitoring of control / comparison groups (in a trial context), context / moderators, monitoring of control groups (in a trial context)
	Questionnaires: endline survey	Online	Nominated teachers, TAs, and headteachers in treatment schools	Descriptive statistics Thematic analysis	2, 4, 5, 7, 9, 10, 14, 15, 16	Moderators, mediators, adaptation, fidelity, context, programme differentiation, quality, unintended consequences
Key informant interviews	Semi-structured interview protocol	In person	Headteachers, teachers, and TAs in focus schools (12)	Thematic analysis Realist evaluation	4, 6, 7, 10, 11, 13	Context / moderators, mediators, fidelity, quality, participant responsiveness
Focus group discussions	Focus group discussion protocol	In person	Headteachers, teachers, and TAs in focus schools (12)	Thematic analysis Realist evaluation	4, 6, 7, 10, 11, 13	Context / moderators, mediators, fidelity, quality, participant responsiveness

Observations	Training observation protocol	Online	Headteachers, teachers, and TAs in focus schools (12)	Thematic analysis Realist evaluation	3, 10	Fidelity, dosage, moderators, mediators, quality
	Structured session observation protocol	In person	Focus schools (12)	Thematic analysis Realist evaluation	3, 8, 9, 12, 14, 16	Fidelity, dosage, monitoring of control groups (in a trial context), programme differentiation, quality, unintended consequences
Document review	Lexia usage data	Digital	Pupils taking part in Lexia	Descriptive statistics Realist evaluation	1, 3	Fidelity, dosage

Practitioner and headteacher surveys

Surveys were the initial mode of data collection and were administered before and after implementing Lexia. This comparative baseline-to-endline approach helped identify changes in practitioners' (teachers and TAs) and headteachers' beliefs and practices resulting from the intervention.

The Lexia coordinator and a nominated Year 2 teacher and/or TA from all participating schools were invited to respond to a baseline survey, which was administered online, before delivery of Lexia. The Lexia coordinator is the person who oversees the more detailed aspects of planning and implementing Lexia: this could be the headteacher in smaller schools, or the literacy coordinator or Year 2 lead; the school's designated Lexia coordinator was agreed by the delivery team with the school in the course of implementing training for teachers, TAs, and school heads. The efficacy trial found TAs to be key programme actors in the delivery of Lexia, hence teachers and TAs were invited to complete the same 'practitioner' surveys.

Surveys consisted primarily of Likert-type questions, with minimal open questions to reduce burden. Questions focused on:

- school context (e.g. available facilities and infrastructure, proportion of FSM-eligible pupils);
- current reading instructional practices, including those specific to struggling readers;
- practitioner knowledge, behaviours, and confidence; and
- school-level policies and structures in support of technology-assisted reading instruction.

After delivering Lexia, nominated practitioners and headteachers were invited to respond to a post-intervention endline survey. Post-intervention practitioner surveys repeated relevant questions from the baseline survey while also asking treatment group practitioners about: i) the relevance and usefulness of training webinars; ii) adjustments made to the intervention (and underlying reasons); iii) fidelity to the programme design; iv) the appropriateness of the pupil selection process and criteria; and v) the impact that Lexia has had on their instructional practices. The post-intervention headteacher survey explored school-wide changes, as well as the financial, material, and human resource requirements of implementing Lexia. Staff in control schools were asked to describe business as usual at baseline and endline, including their IT infrastructure and implementation of potentially similar interventions. It must be noted that since Lexia is intended to be implemented alongside business as usual, this element was also included in the post-intervention endline teacher survey. The rationale behind a baseline and endline approach was to enable a comparison between practitioners' perceptions of the effectiveness of usual practice (i.e. schools' existing reading provisions) and their perceptions of the effectiveness of the Lexia intervention on struggling readers.

Combined, the baseline and endline surveys informed the areas that were selected for further probing through interviews and focus group discussions. To ensure that we captured different school contexts, we selected a subsample of focus schools to act as case studies.

Focus schools

A purposive sampling approach ensured that our analysis accounted for contextual heterogeneity across schools. For each cohort, we focused on six schools from the treatment group and another six from the control group. As such, through the two-year evaluation, we had a total of 24 focus schools evenly split between the treatment and control group. The first 12 focus schools were from the first cohort and the remaining 12 from the second cohort. In both cohorts, six schools each (i.e. three from the control group and three from the treatment group) were selected from EIA and non-EIA contexts.

Schools were purposively sampled based on the following variables:

- urbanicity;
- school size;
- strength of information and communications technology (ICT) infrastructure;⁷ and
- proportion of FSM-eligible pupils.

This sampling scheme allowed for greater contextual depth and provided an opportunity to: i) compare implementation across separate cohorts; ii) explore differences in barriers and facilitators depending on various school characteristics; and iii) describe the extent to which business as usual varies with respect to differences in school context.

The data collection methods used to interrogate focus schools were as follows:

- key informant interviews;
- focus group discussions; and
- observations.

These methods are described in further detail below.

Key informant interviews

Teachers, TAs, and Lexia coordinators from the 24 focus schools participated in key informant interviews. These interviews took place near the end of the intervention and elaborated on survey findings, probing for further detail about: i) how Lexia supplemented other reading instruction and learning activities; ii) the appropriateness of the pupil selection criteria and process; iii) barriers and facilitators to programme implementation; iv) task divisions across school staff; and v) changes in practitioners' knowledge, behaviours, and confidence. Interviews with TAs, teachers, and Lexia coordinators helped to understand any unintended and differential impact of the intervention, particularly on FSM-eligible pupils.

In focus schools from the control group, key informant interviews with practitioners focused on the nature of business as usual, probing into: i) the features of supplementary reading interventions targeted at struggling readers in control schools (if any); ii) how effectively business as usual practices addressed the needs of struggling readers; iii) barriers and facilitators faced by control schools in implementing their respective supplementary reading interventions; and iv) the roles and responsibilities of different staff members in implementing their school's supplementary reading interventions. Key informant interviews with control group schools examined other supplementary reading interventions (as opposed to core, whole-class reading instruction) to ensure like-for-like comparison between Lexia and business as usual.

⁷ Strength of infrastructure was determined through data collected using the baseline survey. Part of the baseline survey asked the school respondent about information relating to ICT infrastructure, such as broadband speed and computer-to-pupil ratio. Schools were classified as having either strong or weak infrastructure.

Furthermore, we interviewed each focus school's headteacher (if this was a different person from the designated Lexia coordinator, in the case of focus schools from the treatment group). While headteachers may not be directly involved in instruction, they can provide high-level information regarding school contexts, including infrastructure, and any changes to business as usual during the year of implementation. Hence, interviews with headteachers centred on school policies, supports, and organisational structures that facilitate the use of technology to aid in reading instruction—and teaching practice more generally—within their respective school.

We use key informant interviews to add further depth to the higher-level data collected through the baseline and endline surveys, and interviewing both headteachers and teaching staff provided another opportunity for data verification. For instance, we explored if the policies and structures reported by headteachers are indeed in place on the ground. In sum, the key informant interviews provided further qualitative depth regarding contexts and mechanisms that lead to Lexia's target outcomes.

Training observation

Guided by a semi-structured observation protocol, we observed recordings of the training webinars delivered by Lexia to headteachers, teachers, and TAs in the 12 focus schools from the treatment group (i.e. six schools in each cohort). The observation protocol was based on the training curriculum provided by the delivery team.

Training observations explored how training content and delivery compared to the original design and adhered to prescribed content, as well as the degree to which headteachers, teachers, and TAs engaged with the training. In sum, training observations primarily collected data on training fidelity.

Classroom observations and setting inventory

Classroom observations were guided by a semi-structured protocol. We conducted observations across the 24 focus schools, which took place in the middle of programme implementation.

In focus schools from the treatment group, these observations examined teaching practices and pupil engagement with Lexia. Furthermore, we observed for implementation fidelity, noting divergences from the design and the corresponding reasons. In addition, we also observed for instances of scaffolding, noting what this entails, how teachers accomplish this, and how Lexia supports this practice.

On the other hand, classroom observations in focus schools from the control group explored the features of other supplementary reading interventions, the extent to which these features promoted pupil engagement and learning, and how these other interventions compared with Lexia sessions.

To guide the classroom observations, we used a modified version of the Reading Lesson Observation Framework (RLOF) (Henk *et al.*, 2000). While the RLOF only allows the observer to assign one of four ratings to each component covered by its inventory of practices, we provided the observer a space to add qualitative comments. Allowing these qualitative comments gave us a clearer picture of instructional practices in focus schools from both the treatment and control group.

A semi-structured observation instrument was used to conduct mid-intervention observations. This instrument is partly adapted from the RLOF (Henk *et al.*, 2000) and focuses on the following elements: i) classroom climate; ii) accessing Lexia; iii) independent learning; and iv) guided reading. Several pedagogical elements captured by the RLOF were omitted from the Lexia session observation instrument since many of those pedagogical elements will occur within the software; the Lexia session observations instead account for in-person practices that take place outside the software.

The observation instrument used the same scoring system as the RLOF (i.e. Observed, Commendation, Recommendation, Not applicable [OCRN]), with practices scored into one of the following categories (Henk *et al.*, 2000), where relevant:

- **Observed.** When the practice was observed and judged to be of satisfactory quality.

- **Commendation.** When the practice was observed and judged to be of very high quality.
- **Recommendation.** When the practice was not observed or was judged to be of unsatisfactory quality.
- **Not applicable.** When the practice was not observed but only because it was not appropriate to the lesson at hand.

A key difference of the Lexia session observation instrument from the RLOF is the supplementation of the OCRN scoring with qualitative comments. This addition of qualitative comments allows for a more in-depth understanding of Lexia implementation and makes the RLOF model more fit for purpose. While the RLOF was designed to ensure consistency in and improve the delivery of reading instruction across schools (Henk *et al.*, 2000)—thus making a closed and structured instrument appropriate—the Lexia observation instrument was used to understand fidelity to (and deviation from) the programme design. Qualitative comments provide a clearer view of these elements.

It is important to note that Lexia’s instructional design inherently uses scaffolding in that it adjusts learning content based on pupils level of content mastery. As such, the IPE focuses on scaffolding performed by teachers *in addition to* the scaffolded instruction already implemented through Lexia. Additional scaffolding that was observed in the IPE therefore pertains to supplementary learning activities done outside Lexia to support computer-led instruction.

In addition, we specifically observed for instances of additional scaffolding provided to FSM, SEND, and EAL pupils. Since we include FSM, SEND, and EAL status as subgroup variables in the impact evaluation, Lexia session observations supplement the impact evaluation’s findings, allowing us to not only quantify Lexia’s differential impact on FSM, SEND, and EAL pupils but also to identify possible explanations underlying that differential impact.

Alongside classroom observations, we conducted an infrastructural inventory of the focus schools. Through the setting inventory, we made note of the different facilities and resources available to practitioners to support the implementation of Lexia. Data collected through the setting inventory was checked against the available infrastructure reported in the headteacher survey. This allowed us to document the available infrastructure within focus schools, providing another opportunity to harmonise findings.

Taken together, the classroom observations and setting inventory allowed us to not only verify self-reported survey data and the information provided through the key informant interviews, but also to probe into the findings of the impact evaluation.

Analysis

The IPE data analysis was conducted through a combination of quantitative analysis and qualitative analytical techniques using the data collected from the research methods described above (see ‘Research methods’ subsection above). This allowed for detailed comparison of treatment and control schools to understand the causal pathways that determine the impact of the Lexia intervention on struggling readers, compared to business as usual.

Descriptive statistics (e.g. mean, range, frequency) were used to analyse and present quantitative data from the practitioner and headteacher surveys, training attendance logs, software use data, and training and session observations. This provides insight into the extent to which different views are representative of the population involved in the trial, compliance with required training, and differences in contexts, beliefs, and practices across participating schools.

All descriptive and inferential statistical analyses were performed using Stata.

Qualitative data were analysed using a thematic analysis approach (Braun and Clarke, 2006). Data from key informant interviews (transcripts from audio recordings and notes made during interviews), responses to open-ended questions from the practitioner surveys, and training and reading session observations were analysed inductively, based on relevant IPE RQs. Analysts first familiarised themselves with the data and coded small subsets to identify initial codes inductively, meaning that this coding was driven by content from responses, rather than predetermined codes. They then convened to

identify overlapping and differing codes, and these initial codes were refined accordingly. Further coding made use of these developed codes while allowing for new codes to emerge.

The themes that surfaced were also analysed using techniques from realist evaluation (Pawson and Tilley, 1997). This helped to identify how different contexts and mechanisms interacted to produce different outcomes across schools. Realist evaluation techniques accounted for contextual heterogeneity across focus schools, which allowed the evaluation team to explore differences in barriers and facilitators that may be contingent on differences in school characteristics.

Costs

The cost evaluation of Lexia was based on the unit price of each software use licence. Using individual licence costs, we computed a per-pupil-school-year cost for implementing Lexia, enabling comparison with other programmes, as required by the EEF's cost evaluation guidance (EEF, 2023). However, we did not exclude the possibility of uncovering any hidden savings and costs from implementing Lexia (e.g. time spent away from teaching to participate in training, supply teacher costs, purchase of supplementary instructional materials, time saved on correcting pupil assessments, time saved on developing differentiated lesson plans, and cost savings from decreased use of printed instructional and assessment materials), which may have surfaced through surveys and interviews with headteachers and practitioners. Where hidden savings and costs did exist, they were monetised using market estimates, which were factored into our computation of a per-pupil-school-year cost.

The Lexia licence cost was compared to the cost of business as usual in control schools; business as usual cost data was collected through surveys with headteachers. Business as usual costs encompassed both direct and indirect costs associated with similar programmes targeting struggling readers. Examples of the direct and indirect costs of business as usual included but were not limited to teaching time rerouted towards training, prices of instructional materials, and additional staff workload required to deliver other programmes targeting struggling readers.

Timeline

Table 7: Timeline

Dates	Activity	Staff responsible / leading
February 2023 – July 2023	Recruitment of schools (Cohort 1)	Queen's University of Belfast
June 2023 – July 2023	Baseline assessment (Cohort 1)	Elklan
July 2023	Randomisation of schools into control and treatment groups (Cohort 1)	RAND Europe
	Informing schools and LexiaUK of randomisation allocations (Cohort 1)	RAND Europe
	Setting up Lexia licences (Cohort 1)	LexiaUK
September 2023 – November 2023	Delivery of Lexia training sessions 1 and 2 (Cohort 1)	LexiaUK
	Administration of baseline surveys to headteachers and teachers/TAs (Cohort 1)	RAND Europe
	Training observations, sessions 1 and 2 (Cohort 1) ^a	
October 2023 – May 2024	Delivery of Lexia training sessions 3 and 4 (Cohort 1)	LexiaUK
	Training observations, sessions 3 and 4 (Cohort 1) ^a	RAND Europe
September 2023 – May 2024	Delivery of Lexia with treatment schools (Cohort 1)	LexiaUK and RAND Europe
February 2024 – April 2024	Midline classroom observations and focus school visits (Cohort 1)	RAND Europe
May 2024 – June 2024	Endline assessment (Cohort 1)	Elklan

Dates	Activity	Staff responsible / leading
	Administration of endline surveys to headteachers and teachers/TAs (Cohort 1)	RAND Europe
	Interviews and focus group discussions with headteachers and teachers/TAs from focus schools (Cohort 1)	
October 2023 – July 2024	Recruitment of schools (Cohort 2)	Queen's University of Belfast
June 2024 – July 2024	Baseline assessment (Cohort 2)	Elklan
July 2024	Randomisation of schools into control and treatment groups (Cohort 2)	RAND Europe
	Informing schools and LexiaUK of randomisation allocations (Cohort 2)	RAND Europe
	Setting up Lexia licences (Cohort 2)	LexiaUK
September 2024 – November 2024	Delivery of Lexia training sessions 1 and 2 (Cohort 2)	LexiaUK
	Administration of baseline surveys to headteachers and teachers/TAs (Cohort 2)	RAND Europe
	Training observations, sessions 1 and 2 (Cohort 2)	
October 2024 – May 2025	Delivery of Lexia training sessions 3 and 4 (Cohort 2)	LexiaUK
	Training observations, sessions 3 and 4 (Cohort 2)	RAND Europe
September 2024 – May 2025	Delivery of Lexia with treatment schools (Cohort 2)	LexiaUK RAND Europe
February 2025 – April 2025	Midline classroom observations and focus school visits (Cohort 2)	RAND Europe
May 2025 – June 2025	Endline assessment (Cohort 2)	Elklan
	Administration of endline surveys to headteachers and teachers/TAs (Cohort 2)	RAND Europe
	Interviews and focus group discussions with headteachers and teachers/TAs from focus schools (Cohort 2)	
June 2025 – November 2025	Drafting of final evaluation report	RAND Europe

^a Since training sessions will be video recorded, the training observations can take place at a later date.

Impact evaluation results

Participant flow including losses and exclusions

Figure 2 below presents the number of schools and pupils involved in the trial at the recruitment, allocation, follow-up, and analysis stages. Note that, while 13,740 schools were approached directly at the recruitment stage, other schools may also have received an invitation, for instance through their local authority or Multi-Academy Trust (MAT).

Figure 2: Participant flow diagram (two arms)

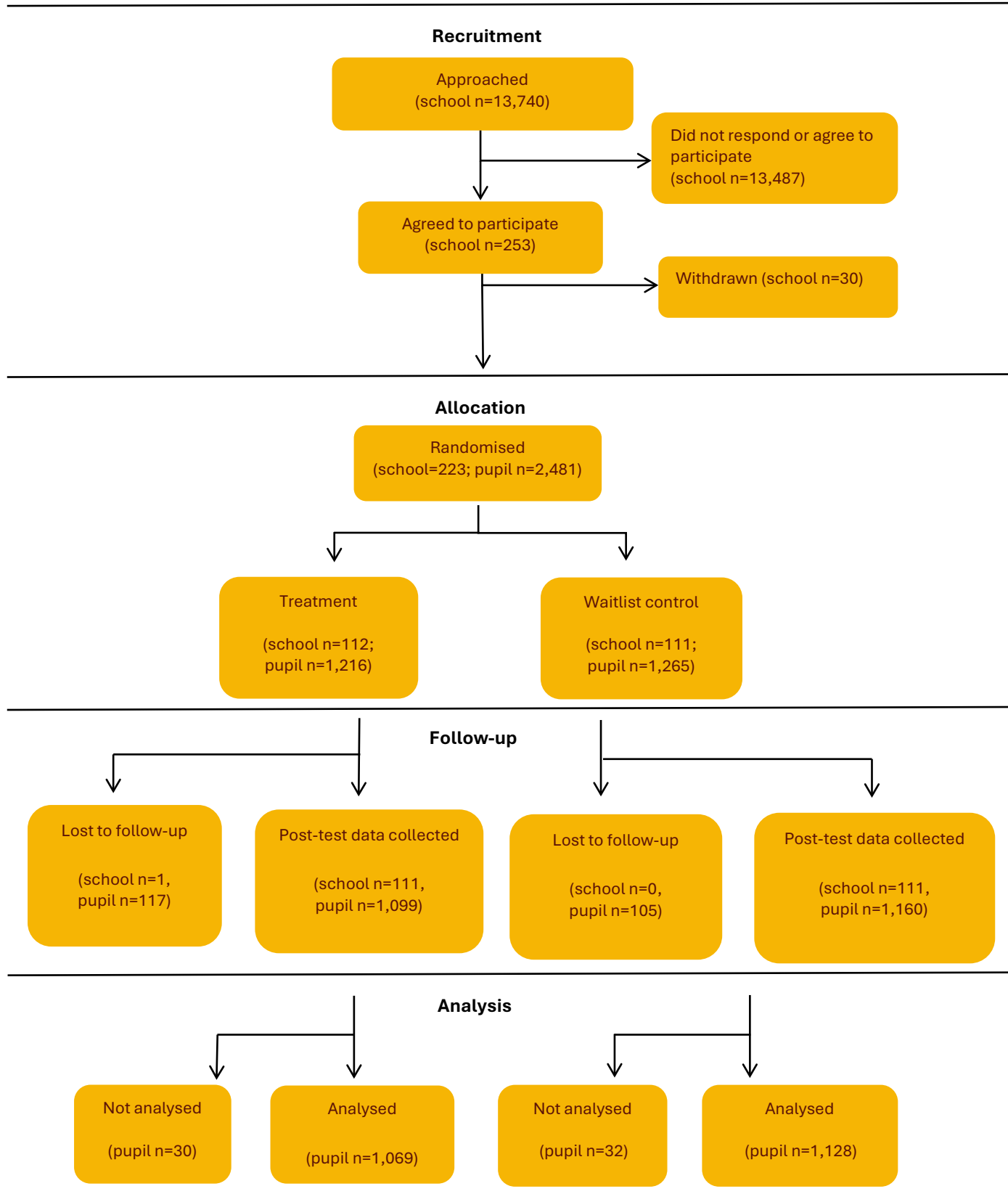


Table 8: MDES at different stages

		Protocol		Randomisation		Analysis	
		Overall	FSM	Overall	FSM	Overall	FSM
MDES		0.115	0.159	0.116	0.160	0.086	0.138
Pre-/post-test correlations	Level 1 (pupil)	0.75	0.74	0.75	0.74	0.755	0.769
	Level 3 (school)	0	0	0	0	0.913	0.875
ICCs	Level 3 (school)	0.06	0.075	0.06	0.075	0.071	0.074
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-sided	Two-sided	Two-sided	Two-sided	Two-sided	Two-sided
Average cluster size		12	4	12	4	9.896	3.261
No. of schools	Intervention	112	112	111	111	111	111
	Control	112	112	111	111	111	111
	Total:	224	224	222	222	222	222
No. of pupils	Intervention	1,344	448	1,221	444	1,069	364
	Control	1,344	448	1,221	444	1,128	360
	Total:	2,688	896	2,442	888	2,197	724

Table 8 shows the MDES at different stages of the trial. At the analysis stage we find an MDES of 0.086, meaning that our trial was powered to detect effect sizes greater than 0.086. This is lower than anticipated at the protocol and randomisation stages. The MDES for the FSM subgroup analysis is likewise lower than anticipated at 0.138.

Attrition

As displayed in Table 9, attrition was over 10% in both the treatment and control cohorts, with 13.75% dropout in the intervention group, 12.15% in the control group and 12.93% overall. One school that was assigned to treatment in Cohort 1 dropped out due to a change in leadership. The remaining attrition was at the pupil level (i.e. spread across schools). Pupils lost to follow-up either left their school or were unavailable on test day. In some cases, pupils did not complete all of the subtests of the composite, meaning that they were not included in the primary outcome.

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

		Intervention	Control	Total
No. of pupils	Randomised	1,216	1,265	2,481
	Analysed	1,069	1,128	2,197
Pupil attrition (from randomisation to analysis)	Number	147	137	284
	Percentage	13.75%	12.15%	12.93%

Pupil and school characteristics

Table 10 shows child and school characteristics as randomised. Overall, there were more schools in EIAs than outside of EIAs, and these were evenly distributed across treatment and control. At the child level, there was balance in baseline attainment between the intervention and control groups, there were similar numbers of pupils across different thresholds of attainment, and the FSM-eligible, EAL, and SEND characteristics. Table 10 shows the characteristics in the analysed sample, where this balance is maintained. One school-level variable that stands out is that there is a higher proportion of schools with 'Outstanding' headline Office for Standards in Education, Children's Services and Skills (Ofsted) grades in the treatment group (12.61%) compared to the control group (5.41%). While the numerical difference is small, this could potentially pose a risk if the progress in pupils in 'Outstanding' schools is faster than that in other schools.

Distributions of the pre-test outcomes are presented as histograms in Appendix E. Distributions of the WRMT-III composite measure are similar in the control and intervention groups and approximately normally distributed. Distributions for the different subtests appear less similar at baseline; however, as shown in Table 11, the mean and SD for the two groups are very close. The distributions for individual subtests do not have the same smooth appearance as those of the composite measure, due to the conversion tables from raw to standardised marks meaning that some values are unrepresented in the scales of subtest standardised marks.

Table 10: Baseline characteristics of groups as randomised

School level (categorical)	National-level mean	Intervention group		Control group	
		n/N (missing)	Count %	n/N (missing)	Count %
EIA		59/112 (0)	52.68%	58/111 (0)	52.25%
Non-EIA		53/112 (0)	47.32%	53/111 (0)	47.75%
Urban		93/112 (0)	83.04%	91/111 (0)	81.98%
Rural		29/112 (0)	16.96%	30/111 (0)	18.02%
'Outstanding' Ofsted rating ^a		14/111 (1)	12.61%	6/111 (0)	5.41%
'Good' Ofsted rating		89/111 (1)	80.18%	95/111 (0)	85.59%
'Requires improvement' Ofsted rating		6/111 (1)	5.41%	7/111 (0)	6.31%
'Inadequate' Ofsted rating		2/111 (1)	1.80%	3/111 (0)	2.70%
School level (continuous)		n/N (missing)	% (SD)	n/N (missing)	% (SD)
Proportion of FSM-eligible pupils		109/112 (3)	27.9 (0.168)	111/111 (0)	24.8 (0.147)
Pupil level (categorical)		n/N (missing)	Count %	n/N (missing)	Count %
FSM eligible		453/1,290 (9)	35.12%	466/1,384 (7)	33.67%
EAL		281/1,289 (10)	21.80%	307/1,383 (8)	22.20%
SEND		364/1,299	28.02%	390/1,391	28.04%

		(0)		(0)		
WRMT-III (Bottom 10%)		9/1,216 (83)	0.74%	12/1,265 (128)	0.95%	
WRMT-III (Bottom 20%)		55/1,216 (83)	4.52%	39/1,265 (128)	3.08%	
WRMT-III (Bottom 30%)		151/1,216 (83)	12.42%	134/1,265 (128)	10.59%	
WRMT-III (Bottom 40%)		329/1,216 (83)	27.06%	337/1,265 (128)	26.64%	
WRMT-III (Bottom 50%)		655/1,216 (83)	53.60%	678/1,265 (128)	53.87%	
WRMT-III (Top 50%)		561/1,216 (83)	46.13%	587/1,265 (128)	46.40%	
WRMT-III (Top 40%)		286/1,216 (83)	23.52%	287/1,265 (128)	22.69%	
WRMT-III (Top 30%)		94/1,216 (83)	7.73%	96/1,265 (128)	7.59%	
WRMT-III (Top 20%)		15/1,216 (83)	1.23%	22/1,265 (128)	1.74%	
WRMT-III (Top 10%)		1/1,216 (83)	0.08%	1/1,265 (128)	0.08%	
Pupil level (continuous)		n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)	Effect size
WRMT-III: Composite score		1,216/1,299 (83)	332.09 (34.16)	1,265/1,393 (128)	322.58 (33.10)	-0.014
Word Identification score		1,236/1,299 (63)	105.79 (13.38)	1,281/1,393 (112)	105.86 (12.81)	-0.005
Word Attack score		1,223/1,299 (76)	113.39 (12.86)	1,276/1,393 (117)	113.92 (12.32)	-0.043
Passage Comprehension score		1,220/1,299 (79)	102.69 (11.46)	1,270/1,393 (123)	102.60 (11.75)	0.008

^a Where the most recent Ofsted rating was conducted after the inspection reform in September 2024 which removed headline grades, these were taken from the last report before the report.

Table 11: Baseline characteristics of groups as analysed

School level (categorical)	National-level mean	Intervention group		Control group		
		n/N (missing)	Count %	n/N (missing)	Count %	
EIA		58/111 (0)	52.25%	58/111 (0)	52.25%	
Non-EIA		53/111 (0)	47.75%	53/111 (0)	47.75%	
Urban		93/111 (0)	83.78%	91/111 (0)	81.98%	
Rural		28/111 (0)	16.22%	20/111 (0)	18.02%	
'Outstanding' Ofsted rating		14/110 (1)	12.73%	6/111 (0)	5.41%	
'Good' Ofsted rating		89/110 (1)	80.91%	95/111 (0)	85.59%	
'Requires improvement' Ofsted rating		6/110 (1)	5.45%	7/111 (0)	6.31%	
'Inadequate' Ofsted rating		1/110 (1)	0.90%	3/111 (0)	2.70%	
School level (continuous)		n/N (missing)	% (SD)	n/N (missing)	% (SD)	
Proportion of FSM- eligible pupils		108/111 (3)	27.9 (0.168)	111 (0)	24.8 (0.147)	
Pupil level (categorical)		n/N (missing)	Count %	n/N (missing)	Count %	
FSM eligible		364/1,069 (130)	34.05%	360/1,126 (265)	31.97%	
EAL		230/1,068 (131)	21.54%	254/1,125 (266)	22.58%	
SEND		291/1,069 (130)	27.22%	318/1,126 (265)	28.24%	
WRMT-III (Bottom 10%)		6/1,069 (230)	0.56%	6/1,128 (265)	0.53%	
WRMT-III (Bottom 20%)		44/1,069 (230)	4.12%	27/1,128 (265)	2.39%	
WRMT-III (Bottom 30%)		127/1,069 (230)	11.88%	104/1,128 (265)	9.21%	
WRMT-III (Bottom 40%)		277/1,069 (230)	25.91%	284/1,128 (265)	25.18%	
WRMT-III (Bottom 50%)		558/1,069 (230)	52.20%	583/1,128 (265)	51.68%	
WRMT-III (Top 50%)		511/1,069 (230)	47.80%	545/1,128 (265)	48.32%	

WRMT-III (Top 40%)		259/1,069 (230)	24.23%	274/1,128 (265)	24.29%	
WRMT-III (Top 30%)		85/1,069 (230)	7.95%	93/1,128 (265)	8.24%	
WRMT-III (Top 20%)		15/1,069 (230)	1.40%	22/1,128 (265)	1.95%	
WRMT-III (Top 10%)		1/1,069 (230)	0.09%	1/1,128 (265)	0.09%	
Pupil level (continuous)		n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)	Effect size
WRMT-III Composite score		1,069/1,299 (230)	323.13 (33.91)	1,128/1,393 (265)	324.46 (32.40)	-0.040
Word Identification score		1,092/1,299 (207)	106.22 (13.26)	1,146/1,393 (247)	106.53 (12.58)	-0.024
Word Attack score		1,081/1,299 (218)	113.67 (12.79)	1,145/1,393 (248)	114.57 (12.07)	-0.072
Passage Comprehension score		1,079/1,299 (220)	102.98 (11.46)	1,138/1,393 (255)	103.17 (11.53)	-0.017

Outcomes and analysis

Primary analysis

As presented in the theory of change, the primary goal of Lexia is to improve reading attainment among struggling readers. As described in the 'Outcome measures' subsection in the 'Methods' section above, reading attainment was captured by a composite measure of three subtests (Word Identification, Word Attack, and Passage Comprehension) within the WRMT-III suite of measures, capturing basic reading skills and ability to understand passages of text. At baseline, the mean composite score was 323.8 with an SD of 33.1 and at endline, the mean of the composite was 306.8 with an SD of 36.3. The histograms in Appendix E suggest that the distribution of scores is approximately normal at baseline and at endline. We discuss the implications of the decrease in average scores between baseline and endline below.

The impact of Lexia on this composite measure of reading attainment was explored using a linear multilevel model with random intercepts, outlined above in the 'Statistical analysis' subsection in the 'Methods' section above. This accounts for the multilevel structure of the data where pupils (the unit of measurement) are nested within schools (the unit of randomisation). The composite WRMT-III measure is taken as the primary outcome measure. The diagnostic plots in Appendix F show that the residuals of this model are not normally distributed. Therefore, the p-values and CIs of the estimated treatment effect were bootstrapped to correct for the violation of this assumption of the model.

Table 12 shows the results of the primary analysis, examining the impact of being assigned to the intervention group on overall reading attainment. The estimated effect size is 0.145, with a 95% CI of 0.075–0.213. We are therefore able to reject the null hypothesis of no effect. In other words, Lexia has a statistically significant effect on reading attainment as measured by the WRMT-III composite. The ICC of this model is 0.071, as reported in Table 5 and Table 8 above.

The EEF guidance (EEF, 2021) suggests an estimate of this size (0.145) can be interpreted as being equivalent to two months' progress.⁸ The CI can similarly be converted to between one and three months' progress. This provides evidence that Lexia is likely to lead to an improvement in reading skills, a short-term outcome of the theory of change.

⁸ The EEF guidance suggests a 1 SD improvement in test scores is equivalent to 12 months' progress (EEF, 2021) see: 'Teaching and Learning/Early Years Toolkit Guide'.

Though it should be noted that the mean composite test scores declined between the pre- and post-tests in both the intervention (from 323.1 to 308.9) and control (from 324.5 to 304.7) groups. However, following the WRMT-III guidance test scores are age standardised to reflect the expected attainment of pupils in the same age group. While these scores decline, the raw scores underpinning these age-standardised scores increase (see Table 13). This suggests that pupils in both the treatment and control group make absolute progress in reading, but less than would be expected by the test. This could be due to the test having unreasonable expectations of progress at this age, or it could be due to our sample not being representative of all pupils given the intervention is targeted at low prior attainers. In either case, we observe the pupils in the treatment group making relatively more progress than pupils in the control group, and this indicates that the intervention has a positive effect.

Table 12: Primary analysis

Outcome	Unadjusted means				Effect size		
	Intervention group		Control group		Total n (intervention; control)	Hedges' g (95% CI)	P-value
	N (missing)	Mean (95% CI)	N (missing)	Mean (95% CI)			
WRMT-III Composite	1,069 (18)	308.89 (306.62, 311.16)	1,128 (16)	304.73 (302.70, 306.76)	2,197 (1,069; 1,128)	0.145 (0.075, 0.213)	<0.001

Table 13: Change in raw scores

Study group	WRMT-III subtests	Baseline				Endline						Difference
		Obs.	Mean	SD	Min.	Max.	Obs.	Mean	SD	Min.	Max.	
Control	Word Identification	1,281	10.18	5.54	0	32	1,156	18.00	6.35	0	36	+7.82
	Word Attack	1,276	7.56	4.41	0	22	1,159	10.66	4.90	0	25	+3.10
	Passage Comprehension	1,270	5.70	3.58	0	16	1,156	9.72	3.70	0	22	+4.01
Treatment	Word Identification	1,236	10.34	5.75	0	30	1,096	18.52	6.86	0	37	+8.18
	Word Attack	1,223	7.47	4.47	0	22	1,095	11.47	5.38	0	25	+4.00
	Passage Comprehension	1,220	5.83	3.48	0	15	1,096	10.23	3.77	0	23	+4.40

Max.=maximum; Min.=minimum; Obs.=observation.

Secondary analysis

The different subtests used in the composite (Word Identification, Word Attack, and Passage Comprehension) and the Oral Reading Fluency subtest, capture different elements of reading proficiency:

- **Word Identification subtest.** Tests the ability to read and pronounce words.
- **Word Attack subtest.** Tests the ability to apply phonics.
- **Passage Comprehension subtest.** Tests the reading comprehension.
- **Oral Reading Fluency subtest.** Captures the ability to read accurately and fluently.

These are all specific reading skills targeted by Lexia according to its theory of change. The secondary analysis examines if there is differential progress across different reading skills as captured by these subtests. Four secondary models were run, one for each subtest. The models for the Word Identification, Word Attack, and Passage Comprehension subtests use the

same model structure as used in the primary analysis, taking the endline score of the subtest as the outcome measure, and adjusting for the subset score at baseline. A lack of data at baseline for the Oral Reading Fluency subtest means that we use the composite of the other three subtests as the baseline measure, with Oral Reading Fluency score at endline used as the outcome measure. As in the primary analysis, the CIs of this analysis were bootstrapped to account for non-normality of the model residuals.

Results of these four models are shown in Table 14. Effect size estimates are statistically significant across the Word Identification, Word Attack, and Passage Comprehension tests, providing evidence that Lexia likely leads to progress in all these domains. The subtest with the lowest effect size estimate is the Word Identification test (0.079), and the Word Attack test has the highest estimate (0.175). Both of these tests are intended by the developer to capture similar basic reading skills, so this likely reflects differences in the tests themselves rather than any underlying difference in skills. The effect size estimate for the Passage Comprehension test for reading comprehension is 0.133. All these estimates are within the 95% CI of the primary effect on the composite score.

The estimated effect size of Lexia on the Oral Reading Fluency subtest is smaller than for the other subtests (0.025). The bootstrapped CI crosses zero and ranges from -0.065 to 0.112. With a p-value of 0.601, there is insufficient evidence to support the hypothesis that Lexia has a non-zero effect size on Oral Reading Fluency. We therefore, find no evidence that Lexia had an effect on Oral Reading Fluency. However, given the difficulty of administering this subtest, only pupils who were more capable readers were able to complete it, risking the introduction of bias into these results.

Table 14: Secondary analysis

Outcome	Unadjusted means				Effect size		
	Intervention group		Control group		Total n (intervention; control)	Hedges' g (95% CI)	P-value
	N (missing)	Mean (95% CI)	N (missing)	Mean (95% CI)			
Word Identification	1,092 (1)	103.55 (102.60, 104.50)	1,146 (2)	102.58 (101.72, 103.43)	2,238 (1,092; 1,146)	0.079 (0.015, 0.142)	0.020
Word Attack	1,081 (11)	107.18 (106.35, 108.01)	1,145 (6)	105.47 (104.75, 106.19)	2,226 (1,081; 1,145)	0.175 (0.089, 0.259)	<0.001
Passage Comprehension	1,079 (14)	98.04 (97.31, 98.77)	1,138 (11)	96.57 (95.87, 97.28)	2,217 (1,079; 1,138)	0.133 (0.043, 0.217)	0.004
Oral Reading Fluency	972 (14)	100.09 (99.17, 101.00)	1,024 (13)	100.00 (99.16, 100.84)	1,996 (972; 1,024)	0.025 (-0.065, 0.112)	0.601

Analysis in the presence of non-compliance

The primary analysis above presents the outcome of the Lexia programme on an ITT basis, showing the impact of the intervention as randomised. However, analysis on this basis could be misleading, as the sample could include pupils who were assigned to the intervention group, but who in actuality did not fully receive the treatment. Further analysis was conducted to explore the impact of receiving the intervention on progress in reading. The minimum compliance threshold agreed by the delivery and evaluation teams was that pupils assigned to receive Lexia spend at least 720 minutes using Lexia. This is equivalent to completing at least 60 minutes a week (or two sessions) for a minimum of 12 non-consecutive weeks equal to completing a minimum of 720 minutes over a 24-week period.

Figure 3 shows the distribution of the time spent on Lexia in minutes. An overwhelming majority of pupils in the treatment group met the agreed definition of compliance, with 97.36% of pupils in the sample analysed at endline spending 720 or more minutes on Lexia, and an average usage of 2,687 minutes. Given the extent to which the Lexia was used, we conducted exploratory analysis on how effect sizes changed with different thresholds of 'compliance' equivalent to 2x (1,440 minutes), 3x (2,160 minutes), 4x (2,880 minutes), and 5x (3,600 minutes) the recommended dosage. This was not part of the original analysis plan but motivated by our finding of high compliance.

Figure 3: Distribution of time spent on the Lexia programme in the treatment group (vertical line at 720 minutes)

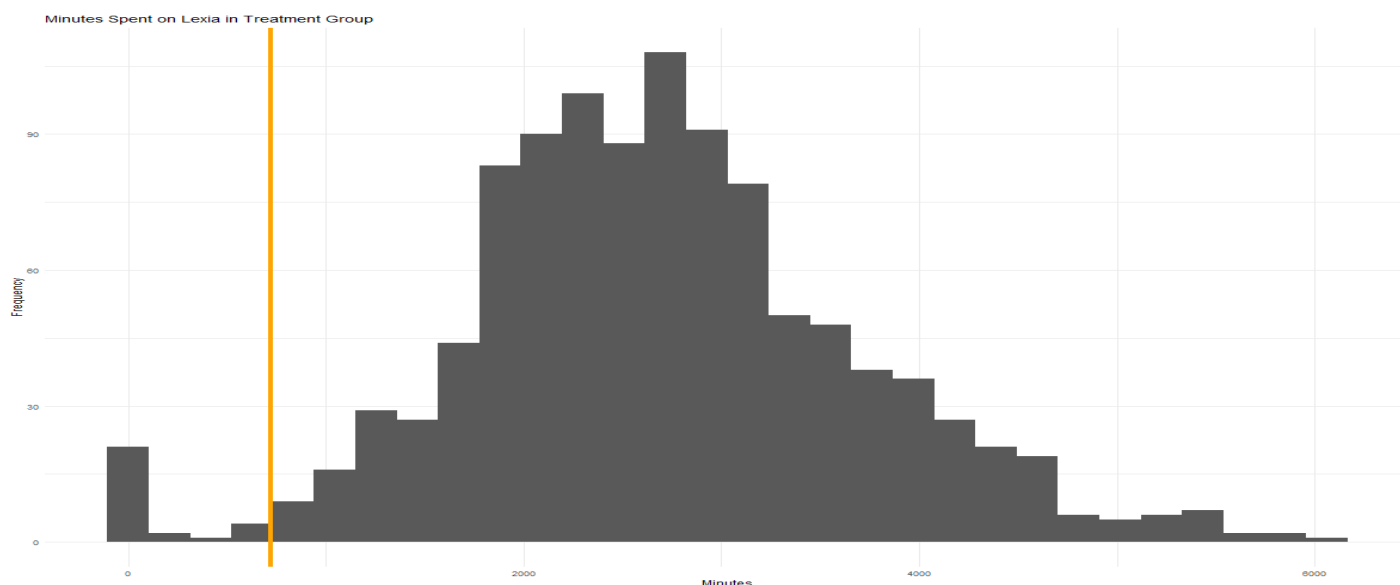


Table 15 shows the results of the CACE analysis described in the ‘Statistical analysis’ subsection in the ‘Methods’ section above, which uses a 2SLS approach. The estimated effect size from the binary compliance model is 0.150 with a 95% CI ranging from 0.081 to 0.219. Unsurprisingly, given the high levels of compliance in the primary sample, this is of a similar magnitude to that of the ITT treatment group in the primary analysis (0.145; Table 12). The dosage model estimates that the average impact of an incremental Lexia session (30 minutes spent on the programme) has a statistically significant effect size of 0.0016.

It does not appear that these impacts diminish with more use of Lexia. Table 15 shows that a pupil receiving 2x the recommended dosage experiences an effect size of 0.158, compared to those receiving less, increasing to 0.202 with 3x the recommended dosage, 0.360 with 4x, and 0.760 with 5x, which is equivalent to ten months’ progress. It is unclear how widely applicable these results are, as selection effects into these different compliance thresholds could bias comparison across the different results. For example, it could be the case that pupils who experience more benefit from Lexia are more likely to spend more time on the programme.

Table 15: Compliance analysis

Outcome	Unadjusted means				Effect size		
	Intervention group		Control group		Total n (intervention; control)	Hedges’ g (95% CI)	P-value
	N (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)			
WRMT-III Composite (compliance analysis)	1,059 (18)	308.95 (306.67, 311.24)	1,128 (16)	304.73 (302.70, 306.76)	2,187 (1,059; 1,128)	0.150 (0.081, 0.219)	<0.001
WRMT-III Composite (dosage analysis)	1,059 (18)	308.95 (306.67, 311.24)	1,128 (16)	304.73 (302.70, 306.76)	2,187 (1,061; 1,133)	0.0016 (0.001, 0.002)	<0.001
WRMT-III Composite (2x Lexias)	1,059 (18)	308.95 (306.67, 311.24)	1,128 (16)	304.73 (302.70, 306.76)	2,187 (1,061; 1,133)	0.158 (0.085, 0.231)	<0.001
WRMT-III Composite (3x Lexias)	1,059 (18)	308.95 (306.67, 311.24)	1,128 (16)	304.73 (302.70, 306.76)	2,187 (1,061; 1,133)	0.202 (0.109, 0.295)	<0.001
WRMT-III Composite (4x Lexias)	1,059 (18)	308.95 (306.67, 311.24)	1,128 (16)	304.73 (302.70, 306.76)	2,187 (1,061; 1,133)	0.360 (0.194, 0.527)	<0.001
WRMT-III Composite (5x Lexias)	1,059 (18)	308.95 (306.67, 311.24)	1,128 (16)	304.73 (302.70, 306.76)	2,187 (1,061; 1,133)	0.760 (0.409, 1.113)	<0.001

Missing data analysis

Attrition in our sample exceeds 5%, which, if non-random, introduces the risk of bias into the results of the primary analysis described above. Following the EEF statistical analysis guidance (EEF, 2022b) this necessitates the exploration of whether there are systematic differences in missingness to see whether the data are missing at random (MAR) using a logistic regression model. We then re-estimate the primary analysis model using multiple imputation.

We conduct a descriptive assessment of the baseline characteristics of the 284 pupils whose data were collected at baseline but were not part of the analysis cohort. For some of these pupils, some of these baseline characteristics were missing. The balance of these pupils with those who were analysed in the sample is reported in Table 16. These results indicate that, on the school level, there are higher rates of missingness in schools with Ofsted ratings 'Requires improvement' and 'Inadequate'. There are also substantial differences in the proportion of FSM, SEND, and EAL pupils in the missing and analysed group. There is also 10 points difference in the baseline score.

Table 16: Comparison of covariates across pupils with missing outcomes and those in the analytical sample

	Attrition group		Analysis group	
	n/N	Count %	n/N	Count %
Treatment	137/284	48.24%	1,254/2,406	52.12%
Control	147/284	51.76%	1,152/2,406	47.88%
EIA	155/284	54.58%	1,289/2,406	53.57%
Non-EIA	129/284	45.42%	1,117/2,406	46.43%
Urban	246/284	86.62%	2,056/2,406	85.45%
Rural	38/284	13.38%	350/2,406	14.55%
'Outstanding' Ofsted rating	20/283	7.07%	207/2,392	8.65%
'Good' Ofsted rating	220/283	77.74%	2,002/2,392	83.7%
'Requires improvement' Ofsted rating	24/283	8.48%	137/2,392	5.73%
'Inadequate' Ofsted rating	19/283	6.71%	46/2,392	1.92%
	Mean	SD	Mean	SD
Proportion of FSM-eligible pupils	0.2858	0.1505	0.2582	0.1553
	n/N	Count %	n/N	Count %
Female	129/284	45.42%	1,131/2,390	47.32%
Male	155/284	54.58%	1,259/2,390	52.68%
FSM	117/271	43.17%	802/2,403	33.37%
EAL	74/271	27.31%	514/2,401	21.41%
SEND	95/271	35.06%	659/2,403	27.42%
Pupil level (continuous)	Mean	SD	Mean	SD
WRMT-III Composite score	310.95	35.13	323.78	33.14

Note: Missingness in some of these variables (Ofsted rating, sex, FSM, EAL, SEND) accounts for the difference in N within columns.

These results give an indication of how characteristics are distributed in the missing and non-missing samples and to strengthen this evidence we constructed a model of missingness using logistic regression to examine whether data is MAR. As discussed in the 'Methods' section we created a dummy indicating missingness and modelling the relationship between

the baseline characteristics above and this variable to examine whether there are any significant predictors of missingness. The results of this regression are presented in Appendix D Table 15.

The results of this regression (Appendix D Table 15) indicate that data could be MAR, with pupils who are FSM eligible ($p = 0.024$) and in Ofsted rated 'Inadequate' schools ($p < 0.001$) being significantly more likely to have data missing. The probability of the pupil's endline data being missing declines significantly with higher scores ($p < 0.001$). EAL pupils are more likely to be missing at endline, but with less statistical certainty with a p -value of 0.098, so this is only significant at the 0.1 level. This model only captures observable variables, and while providing evidence of data being MAR, cannot capture the impact of unobserved characteristics, which could also drive missingness.

Re-estimation of the primary analysis using multiple imputation

The results above suggest that data is not missing completely at random (MCAR), which suggests that the complete case analysis of the primary analysis is biased. Therefore, we use multiple imputation (using MICE methodology) to account for potential MAR data at endline, re-estimating the impact of Lexia on the primary outcome using chained equations which take into account pupil- and school-level effects and the multilevel structure of the data. Our model though cannot capture all the characteristics that determine missingness, and there are likely to be other unobserved factors that are influential. This analysis is limited as it by necessity excludes the pupils who, as well as having missing endline data, are missing the baseline covariates used in the imputation of the primary outcomes. We performed 20 imputations and results were pooled across each imputed dataset. This will also only alleviate bias if data is MAR rather than missing not at random (MNAR).

The results of the primary model, with multiple imputation, are presented in Table 17. The estimated effect size (0.138) is statistically significant and similar to the primary analysis estimate (0.145, Table 12), representing two months' progress. However, it is a slightly smaller effect size providing some evidence that the results of the primary analysis are a slight overestimate. Given the similarity of the results between the multiple imputation analysis and the complete case analysis we have no reason to suspect that the data is MNAR.

Table 17: Missing data analysis results

Outcome	Unadjusted means				Effect size		
	Intervention group		Control group		Total n (intervention; control)	Hedges' g (95% CI)	P-value
	N (missing)	Mean (95% CI)	N (missing)	Mean (95% CI)			
WRMT-III Composite (Imputed)	1,087	305.73 (304.40, 307.06)	1,142	295.20 (293.94, 296.46)	2,229 (1,087; 1,142)	0.138 (0.064, 0.212)	<0.001

Subgroup analyses

FSM eligibility

We examine the impact of Lexia on the FSM subgroup to evaluate whether it has differential effects for pupils from disadvantaged backgrounds. We use two different models in this analysis:

- the first restricts the primary model to the FSM subsample to estimate the effect in that group;
- the second introduces an interaction term between FSM eligibility and treatment to evaluate whether Lexia is a gap-widener or a gap-closer.

The results of the subsample analysis in Table 18 indicate that Lexia has significant positive effects on FSM pupils, with an effect size on WRMT-III composite scores of 0.1707, which is equivalent to two months' progress in reading skills. This is corroborated by the interaction term analysis result in Table 18, which finds a similar effect size of 0.172 (Table 19), although the 95% CIs of this estimate contain a possible zero effect size. This is slightly higher than the effect size on the primary outcome in the overall sample (Table 12). The interaction term coefficient of this model is 1.415 (Table 19); however, the p -

value in Table 19 corresponding to the interaction term for the null hypothesis that there is no difference between FSM pupils and the rest of the sample, indicates that this difference is not statistically significant.

Table 18: FSM subsample analysis

Outcome	Unadjusted means				Effect size		
	Intervention group		Control group		Total n (intervention; control)	Hedges' g (95% CI)	P-value
	N (missing)	Mean (95% CI)	N (missing)	Mean (95% CI)			
WRMT-III Composite	364 (5)	302.7363 (298.6764, 306.7962)	360 (7)	298.0361 (294.4525, 301.619)	724 (384; 360)	0.1707 (0.0847, 0.2567)	<0.001

Table 19: FSM interaction term analysis

Outcome	Unadjusted means				Effect size			
	Intervention group		Control group		Total n (intervention; control)	Hedges' g (for FSM pupils; 95% CI)	Interaction term coefficient	P-value of interaction term coefficient
	N (missing)	Mean (95% CI)	N (missing)	Mean (95% CI)				
WRMT-III Composite	1,069 (18)	308.89 (306.62, 311.16)	1,126 (16)	304.65 (302.62, 306.67)	2,195 (1,069; 1,126)	0.172 (-0.003, 0.348)	1.415	0.490

EAL pupils

The analysis of the impact of Lexia for EAL pupils follows the same models as the FSM analysis. Similarly, the results of the subsample analysis in Table 20 show that the impact of Lexia on EAL pupils' reading is statistically significant. The estimated effect size of 0.224 is equivalent to three months' relative progress (i.e. three fewer months of lost progress). The progress of EAL pupils in the interaction model (Table 21), which acts as a sensitivity check, is still large and suggestive of three months' progress, but marginally lower, and the p-value for the interaction term of 2.069 does not support that progress for EAL pupils is significantly larger than that of non-EAL pupils. Therefore, there is insufficient evidence to suggest that Lexia is a gap-closer for EAL pupils.

Table 20: EAL subsample analysis

Outcome	Unadjusted means				Effect size		
	Intervention group		Control group		Total n (intervention; control)	Hedges' g (95% CI)	P-value
	N (missing)	Mean (95% CI)	N (missing)	Mean (95% CI)			
WRMT-III Composite	230 (4)	314.44 (309.70, 319.18)	254 (1)	303.82 (299.73, 307.90)	484 (230; 254)	0.224 (0.113, 0.335)	<0.001

Table 21: EAL interaction term analysis

Outcome	Unadjusted means				Effect size			
	Intervention group		Control group		Total n (intervention; control)	Hedges' g (for EAL pupils; 95% CI)	Interaction term coefficient	P-value of interaction term coefficient
	N (missing)	Mean (95% CI)	N (missing)	Mean (95% CI)				
WRMT-III Composite	1,068 (19)	308.85 (306.58- 311.11)	1,125 (17)	304.63 (302.60- 306.66)	2,193 (1,068; 1,125)	0.192 (0.009, 0.374)	2.069	0.366

SEND pupils

The estimated impact of Lexia for SEND pupils is low at 0.041 (Table 22) and not statistically significant from zero, which means that we cannot reject the null hypothesis that Lexia does not have an effect on SEND pupils. This marks SEND pupils out from the other subgroups of interest and the primary cohort, whose results are consistent with two months' relative progress in reading ability. The p-value in the interaction term analysis for the null hypothesis that the coefficient on the interaction term is zero is significant at the 0.05 level. This coefficient in Table 23 is negative and has a magnitude of -5.254, suggesting that Lexia is a gap-widener for pupils with SEND.

Table 22: SEND subsample analysis

Outcome	Unadjusted means				Effect size		
	Intervention group		Control group		Total n (intervention; control)	Hedges' g (95% CI)	P-value
	N (missing)	Mean (95% CI)	N (missing)	Mean (95% CI)			
WRMT-III Composite	291 (5)	291.43 (286.65, 296.20)	318 (5)	291.11 (287.05, 295.17)	609 (291; 318)	0.041 (-0.049, 0.131)	0.369

Table 23: SEND interaction term analysis

Outcome	Unadjusted means				Effect size			
	Intervention group		Control group		Total n (intervention; control)	Hedges' g (for SEND pupils; 95% CI)	Interaction term coefficient	P-value of interaction term coefficient
	N (missing)	Mean (95% CI)	N (missing)	Mean (95% CI)				
WRMT-III Composite	1,069 (18)	308.89 (306.62, 311.16)	1,126 (16)	304.65 (302.62, 306.67)	2,195 (1,069; 1,126)	0.039 (-0.144, 0.222)	-5.254	0.017

Baseline reading level

Lexia is a programme designed primarily for struggling readers. It is therefore, interesting to explore whether reading proficiency before the intervention influences the benefit experienced by pupils. To explore this, we divided the sample in strata of baseline attainment and estimated the impact of Lexia on the primary outcome by estimating the primary model in each. Smaller sample sizes in the bottom and top 10% and top 20% of the score distribution precludes performing this analysis for those attainment levels, so we conduct analysis in the bottom 20%, 30%, 40%, and 50% and the top 50%, 40%, and 30% strata.

The results in Table 24 suggest that there is evidence for stronger effect sizes for pupils who scored more highly at baseline. The estimated effect sizes for those who scored within the top 30%, 40%, and 50% of the possible WRMT-III scores were substantially higher than for those in the bottom 50%, and the primary estimate for the overall sample. In these cohorts, we observe statistically significant effect sizes of 0.364, 0.352, and 0.285, respectively, which are equivalent to four to five months' relative progress.

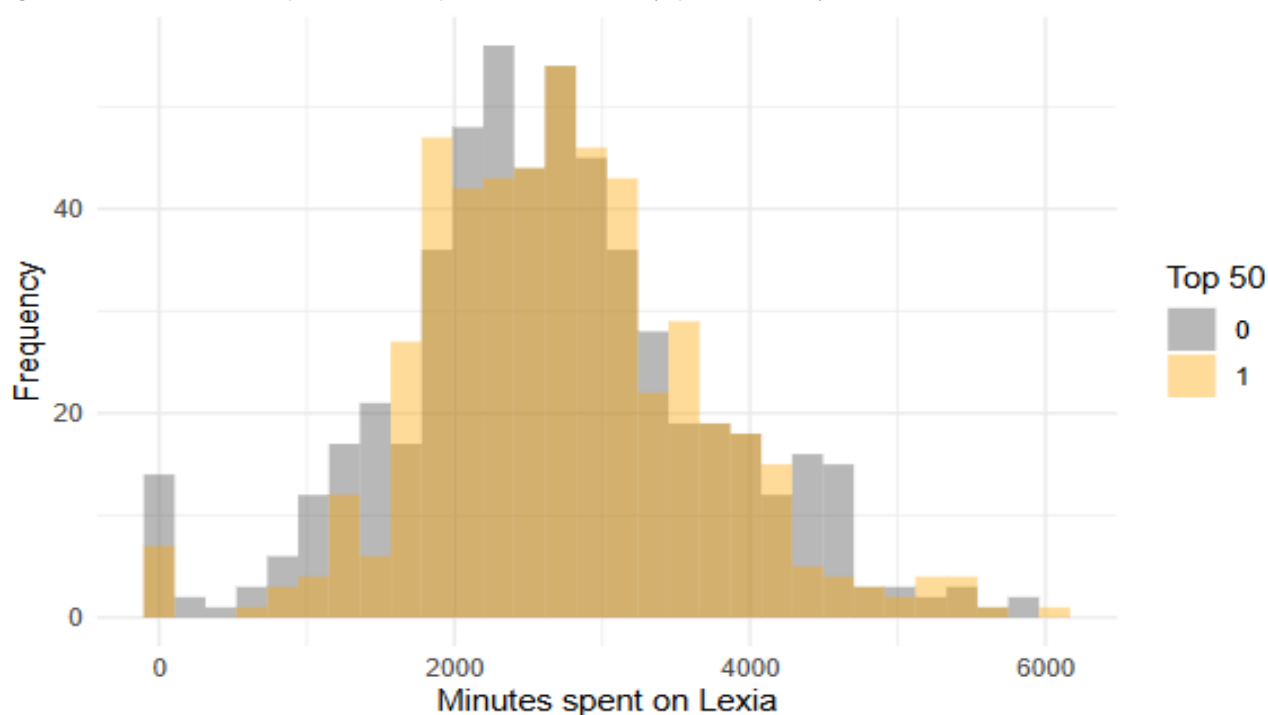
While not statistically significant, the effect sizes estimated for the lower half of attainment (with the exception of the bottom 20%) are still indicative of an average relative progress in reading proficiency of two months (two fewer months of lost progress), for lower ability pupils. Pupils in the bottom 50% of scores at baseline experienced an effect size of 0.118, similar to the effect size of 0.120 for the bottom 40%, and slightly lower than the effect size of 0.199 for the bottom 30% (Table 24). Caution should be exercised in interpreting the latter two of these results, however, as the bootstrapped CIs are wide, and do not allow us to conclude that pupils in these groups did not have a null effect. There is also little evidence for benefit to the bottom 20% of baseline attainment, for whom the estimated effect size is very low at 0.009 and not statistically significant from zero (Table 24). In the context of these results, it is important to note that this part of the evaluation was not sufficiently powered to detect more moderate effect sizes.

Table 24: Subgroup analysis, by baseline reading level cohort

Outcome	Unadjusted means				Effect size		
	Intervention group		Control group		Total n (intervention; control)	Hedges' g (95% CI)	P-value
	N (missing)	Mean (95% CI)	N (missing)	Mean (95% CI)			
WRMT-III Composite (Bottom 20% subgroup)	44 (0)	239.20 (229.98, 248.42)	27 (0)	238.41 (226.57, 250.24)	71 (44; 27)	0.009 (-0.476, 0.456)	0.967
WRMT-III Composite (Bottom 30% subgroup)	127 (0)	254.94 (249.31, 260.57)	104 (0)	250.14 (243.92, 256.37)	231 (127; 104)	0.199 (-0.057, 0.447)	0.129
WRMT-III Composite (Bottom 40% subgroup)	277 (0)	270.37 (266.41, 274.32)	284 (0)	269.90 (266.34, 273.47)	561 (284; 277)	0.120 (-0.043, 0.280)	0.15
WRMT-III Composite (Bottom 50% subgroup)	558 (0)	285.90 (283.21, 288.59)	583 (0)	284.35 (281.86, 286.83)	1,141 (558; 583)	0.118 (0.010, 0.225)	0.032
WRMT-III Composite (Top 50% subgroup)	511 (0)	333.99 (331.79, 336.19)	545 (0)	326.54 (324.52, 328.56)	1,056 (511; 545)	0.285 (0.170, 0.401)	<0.001
WRMT-III Composite (Top 40% subgroup)	259 (0)	346.02 (343.33, 348.71)	274 (0)	338.56 (336.12, 341.00)	533 (259; 274)	0.352 (0.187, 0.515)	<0.001
WRMT-III Composite (Top 30% subgroup)	85 (0)	355.53 (350.83, 360.23)	93 (0)	349.75 (345.37, 354.13)	178 (85; 93)	0.364 (0.085, 0.646)	0.011

In interviews, some practitioners expressed that Lexia was not as beneficial for pupils with lower attention spans, which could be correlated with baseline reading attainment (see the 'Implementation' subsection in the 'IPE results' section below). To explore this, we compared the distributions of minutes spent using Lexia (which was used to measure the intervention dosage). The time spent using the programme was similar for the pupils in the top and bottom 50% of baseline attainment; however, this does not capture the difference in the quality of attention in the time spent using Lexia, or additional support given by teachers. Lack of information on the extent to which pupils actively engaged with Lexia (rather than simply the number of minutes they spent logged into Lexia) is discussed as a limitation of this study in the 'Conclusion' section below.

Figure 4: Distribution of time spent on Lexia, by baseline attainment (top/bottom 50%)



Additional analyses and robustness checks*Cohort analysis*

Recruitment to the Lexia trial took place over two cohorts recruited in different years. Differences in the characteristics of the schools and pupils in these two cohorts, and their impact on the effectiveness of Lexia, could be disguised in the pooled analysis of the primary model. Cohort was included as a factor in the primary model, where its estimated impact on the outcome variable is insignificant, as seen in the regression output in Appendix D Table 1. As an additional sensitivity analysis, we estimated the effect of Lexia on pupils in Cohort 1 and Cohort 2 separately, the results of which are presented in Table 25. The estimate effect size of 0.118 in Cohort 1 and 0.173 in Cohort 2 are both statistically significant and equivalent to two months' relative progress. While the two effect size estimates are not significantly different from each other (they lie in each other's CIs), the difference in magnitude of effect size calculated is still noteworthy.

Table 25: Results by cohort

Outcome	Unadjusted means				Effect size		
	Intervention group		Control group		Total n (intervention; control)	Hedges' g (95% CI)	P-value
	N (missing)	Mean (95% CI)	N (missing)	Mean (95% CI)			
WRMT-III Composite (Cohort 1)	550 (15)	306.45 (303.18, 309.72)	592 (11)	304.54 (301.79, 307.28)	1,142 (550; 592)	0.118 (0.021, 0.213)	0.017
WRMT-III Composite (Cohort 2)	519 (3)	311.47 (308.34, 314.61)	536 (5)	304.95 (301.93, 307.96)	1,055 (519; 536)	0.173 (0.077, 0.269)	<0.001

We explored whether there were underlying differences in the characteristics of the schools and pupils across the two cohorts, in a method similar to the balance tests presented in Table 10 and Table 11. These are presented in Table 26. There are differences in the number of settings in EIAs across the two cohorts, with the first cohort's sample having a higher proportion of EIA schools by 5.4 percentage points. While the levels of baseline attainment are similar in Cohort 1 and Cohort 2, examination of the proportion of pupils in different levels of attainment suggests that the distributions of baseline scores are different, with Cohort 1 having a lower proportion of pupils in the top 50%. Evidence from the subgroup analysis suggests that there are higher effect sizes for pupils with higher attainment at baseline, which could drive the difference in the effect size estimates for each cohort.

Table 26: Cohort characteristic comparisons

School level (categorical)	Cohort 1		Cohort 2	
	n/N (missing)	Count %	n/N (missing)	Count %
EIA	61/111 (0)	54.95%	55/111 (0)	49.55%
Non-EIA	50/111 (0)	45.05%	56/111 (0)	50.45%
Urban	86/111 (0)	79.28%	96/111 (0)	86.49%
Rural	25/111 (0)	10.72%	15/111 (0)	13.51%
'Outstanding' Ofsted rating	8/111 (0)	7.21%	12/110 (1)	10.91%
'Good' Ofsted rating	92/111 (0)	82.88%	92/110 (1)	83.64%

'Requires improvement' Ofsted rating	9/111 (0)	8.11%	4/110 (1)	3.64%	
'Inadequate' Ofsted rating	2/111 (0)	1.80%	2/110 (1)	1.82%	
School level (continuous)	n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)	
Proportion of FSM-eligible pupils	108/111 (3)	0.252 (0.162)	111/111 (0)	0.274 (0.154)	
Pupil level (categorical)	n/N (missing)	Count %	n/N (missing)	Count %	
FSM eligible	382/1,140 (259)	33.51%	342/1,055 (246)	32.42%	
EAL	233/1,139 (260)	20.46%	251/1,054 (247)	23.81%	
SEND	336/1,140 (259)	29.47%	273/1,055 (246)	25.88%	
WRMT-III (Bottom 10%)	8/1,142 (249)	0.70%	4/1,055 (246)	0.38%	
WRMT-III (Bottom 20%)	43/1,142 (249)	3.77%	28/1,055 (246)	2.65%	
WRMT-III (Bottom 30%)	121/1,142 (249)	10.60%	110/1,055 (246)	10.43%	
WRMT-III (Bottom 40%)	320/1,142 (249)	28.02%	241/1,055 (246)	22.84%	
WRMT-III (Bottom 50%)	627/1,142 (249)	54.90%	514/1,055 (246)	48.72%	
WRMT-III (Top 50%)	515/1,142 (249)	45.10%	541/1,055 (246)	51.28%	
WRMT-III (Top 40%)	261/1,142 (249)	22.85%	273/1,055 (246)	25.71%	
WRMT-III (Top 30%)	95/1,142 (249)	8.32%	95/1,055 (246)	7.91%	
WRMT-III (Top 20%)	22/1,142 (249)	1.93%	15/1,055 (246)	1.42%	
WRMT-III (Top 10%)	2/1,142 (249)	0.18%	0/1,055 (246)	0.00%	
Pupil level (continuous)	n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)	Effect size
WRMT-III Composite score	1,142/1,391 (249)	322.53 (33.88)	1,055/1,301 (246)	325.21 (32.29)	-0.081
Word Identification score	1,173/1,391 (218)	105.68 (13.38)	1,065/1,301 (236)	107.15 (12.32)	-0.115
Word Attack score	1,160/1,391 (231)	113.77 (12.22)	1,066/1,301 (235)	114.53 (12.65)	-0.061
Passage Comprehension score	1,153/1,391 (238)	102.69 (11.77)	1,064/1,301 (237)	103.49 (11.20)	-0.070

We combined the two estimates using meta-analysis. The differences above suggest that, despite the simplicity of the process, we should take into account the possibility that the two cohort samples represent two dissimilar underlying populations, in which case one cannot assume that there is a single underlying ‘true’ effect size in both of the two cohorts. We therefore, opted for a random effects model of meta-analysis, which unlike the also common fixed effects approach, does not assume a single effect size. The results are presented in Table 27 below and show that the pooled effect size is very similar to the effect size calculated in the primary analysis, as are the upper and lower bounds. Therefore, this exercise corroborates the finding that the impact of Lexia is statistically significant and equivalent to two months’ progress.

Table 27: Meta-analysis of the cohort estimates

Outcome	Weighted mean effect size	Weighted standard error	Weighted CI low	Weighted CI high
WRMT-III Composite	0.145	0.035	0.078	0.213

Analysis with Oral Reading Fluency

Table 28 summarises the impact of Lexia on the composite of the WRMT-III subtests we used in the primary model, and the Oral Reading Fluency subtest, which was dropped from the primary analysis due to issues with administering the test. However, it was included in the primary outcome of the efficacy trial, so these results offer a more direct comparison. The estimated effect size is 0.134, which is statistically significant with lower and upper bounds suggesting a potential effect size on reading proficiency including oral fluency of 0.056–0.206. As the estimated effect of Lexia on Oral Reading Fluency alone is null, it is unsurprising that including it in the composite lowers the estimate of overall progress. These results should be interpreted with caution, as only pupils who were capable of taking the Oral Reading Fluency subtest at endline, which unlike other subtests, had proven to be too difficult for use at baseline, are included in this analysis, who are presumably more capable readers. Therefore, the presence only of pupils who are more able, risks introducing bias into these results.

Table 28: Analysis of primary outcome plus Oral Reading Fluency

Outcome	Unadjusted means				Effect size		
	Intervention group		Control group		Total n (intervention; control)	Hedges’ g (95% CI)	P-value
	N (missing)	Mean (95% CI)	N (missing)	Mean (95% CI)			
WRMT-III Composite plus Oral Reading Fluency	965 (14)	415.01 (412.12, 417.90)	1,017 (13)	410.15 (407.56, 412.73)	1,982 (965; 1017)	0.134 (0.056, 0.206)	<0.001

Implementation and Process Evaluation results

Findings in this section are based on responses from surveys, interviews, and observations.

Survey responses were as follows: 163 baseline headteachers (73% response rate); 156 baseline practitioners (70% response rate); 141 endline headteachers (63% response rate); and 118 endline practitioners (53% response rate). Practitioners could be either Year 2 teachers or TAs. Responses to endline surveys were split relatively equally between treatment and control schools.

Interviews were conducted with headteachers and practitioners in 24 focus schools, split evenly between treatment and control. Observations were carried out in 23 of these schools, 12 in the treatment group and 11 in the control group.⁹

Key findings

- Fidelity to Lexia was high. Schools were committed to delivering Lexia as intended and experienced few barriers to doing so. Implementation of Lexia was facilitated by high-quality training and support from LexiaUK and by schools' allocation of dedicated equipment, staff, space, and time to Lexia sessions.
- Most schools were already using technology in teaching reading and had interventions in place for struggling readers. However, particularly in schools with resourcing challenges, Lexia's personalised, technology-led approach appears to have enabled more pupils to receive regular instruction at a level appropriate for them.
- Lexia was perceived by teachers to have a positive impact on pupils' reading ability, as well as on the confidence of members of staff with less experience in teaching reading.

Fidelity

This section discusses the following questions:

IPE RQ3: To what extent was the intervention delivered as intended (including the selection of participating pupils, the administration of webinar-style training sessions, and the actual implementation of Lexia)?

IPE RQ10: To what extent did the mandatory training sessions and other available support enable practitioners to deliver the programme?

Lexia was delivered as intended

Data from observations and interviews suggest that schools' delivery of Lexia sessions was almost always as intended by the developers of Lexia. In line with Lexia's theory of change, practitioners were observed to provide good support to pupils during sessions, including behaviour management, help with re-focusing when pupils became distracted, and support with technology where needed. Practitioners were also seen to set out equipment before sessions began to ensure that time spent in Lexia sessions was used efficiently.

Many interviewees emphasised their school's commitment to ensuring that pupils met their targets for time spent on Lexia, even during busy periods for the school. The compliance results (see the 'Outcomes and analysis' subsection in the 'Impact evaluation results' section above) demonstrate that the vast majority of schools were successful in achieving this. Indeed, most pupils exceeded their time targets by a considerable margin, which suggests both that schools made Lexia a priority and that Lexia was relatively straightforward for schools to deliver.

In line with the finding that schools made delivery of Lexia a priority, Lexia sessions in treatment schools often appeared to be delivered more formally and consistently than reading sessions in control schools. For instance, treatment schools were more likely than control schools to set aside a dedicated time and space for pupils to complete sessions in. Treatment

⁹ Due to a miscommunication between the school and the research team, a research team member visited one control group school on a day when no Year 2 reading session was taking place and was therefore unable to carry out an observation.

schools were also more likely than control schools to have established procedures for pupils to follow at the start and end of each session. Again, this appeared to be a reflection of treatment schools' commitment to the trial and desire to ensure Lexia was taken seriously by both teachers and pupils. However, it also suggests that the impact of trial conditions should be taken into account when considering the ecological validity of the trial's results, as it is possible that some of the observed impact of Lexia may stem from the seriousness with which treatment schools approached the trial.

Observations of schools' Lexia sessions suggested that pupils generally appeared to be using Lexia as outlined in the theory of change, i.e. working on different aspects of reading at a level appropriate to their development. In interviews, several practitioners corroborated this finding. However, a small minority of practitioners felt either that pupils were working on material in Lexia that was too simple for them or that pupils were succeeding in completing levels in Lexia without truly mastering the material covered.

Alongside the regular online sessions, Lexia's theory of change also includes the provision of additional one-on-one or whole-group support by practitioners when required. This was not directly observed during the trial, but many practitioners confirmed in interviews that they provided additional support to individual Lexia pupils, generally outside of Lexia group sessions, using the paper-based resources provided by Lexia.

Finally, the theory of change also includes use of Lexia progress reports by practitioners and/or the senior leadership team (SLT). In endline surveys, 67% of practitioners and 78% of headteachers reported using the progress reports generated by Lexia to adjust instruction to pupils' individual needs, suggesting that most schools made active use of these reports. Several interviewees similarly reported finding the data generated by Lexia useful to monitor pupils' progress and pinpoint individual pupils' specific areas of weakness.

Training was well attended and useful

Practitioners and headteachers involved in the trial were invited to attend three required online training sessions with Lexia trainers, as well as one optional online training session. Observations of the training sessions suggest that all key topics were covered as intended in the first three sessions, and that the optional fourth session provided schools with a valuable opportunity to ask questions and review their school's progress in implementing Lexia. Ongoing individualised support was also available from Lexia: schools were assigned a dedicated contact at LexiaUK who could help with troubleshooting and answer any questions.

These training sessions were well attended and received by school staff. At endline, 90% of practitioners reported having had training on the use of Lexia, and 71% had attended all training sessions (including the optional fourth session), suggesting that Lexia training inputs were completed in line with the programme's theory of change. 96% of practitioners and 85% of headteachers felt the sessions had been helpful. In open-text survey responses and interviews, many participants praised the clarity and comprehensiveness of the training as well as the usefulness and responsiveness of the tailored support available from Lexia.

The selection of participating pupils was almost always felt to be appropriate

Data from observations and interviews suggest that the selection of pupils to participate in the trial was appropriate in the vast majority of cases. In interviews, most practitioners felt that the right pupils had been selected; observers concurred with this judgement, noting that most participating pupils appeared able to focus on Lexia for the duration of each session, with some TA support. A small number of practitioners felt that having to choose pupils to participate in the trial when they were in Year 1 had led to certain pupils missing out, particularly those who were not developmentally ready to focus for the duration of a Lexia session in Year 1 but who would have benefited from the intervention in Year 2.

While practitioners were generally very positive about their selection of pupils, several TAs reported that certain pupils, particularly those with SEND or difficulties with concentration, struggled to focus on Lexia or required one-on-one support to use the programme effectively. Data from observations corroborates this: a small minority of pupils were seen to struggle to concentrate on Lexia for more than a couple of minutes at a time, or to need almost constant TA support to remain focused.

Implementation

This section discusses the following questions:

IPE RQ4: What are the barriers and facilitators to the implementation of Lexia? Which barriers and facilitators, if any, were specifically or disproportionately experienced by FSM-eligible pupils? SEND pupils? EAL pupils?

IPE RQ5: What changes or adjustments, if any, were made to the programme during its implementation? Why were these adjustments made?

IPE RQ6: How did settings divide tasks (i.e. facilitating Lexia sessions, weekly progress report monitoring, providing additional support to pupils) among their staff? What are the benefits and disadvantages of the different task divisions observed?

Lexia was generally delivered by TAs and monitored by Year 2 teachers and/or headteachers

In the endline survey, practitioners were asked how their school divided tasks associated with Lexia between staff members. TAs were reported to deliver Lexia sessions by just over three-quarters of practitioners, while 65% of practitioners said that monitoring pupils' progress was the job of the Year 2 teacher. Headteachers were seen to be responsible for managing their school's implementation of Lexia by 55% of practitioners, while an equal number of practitioners stated that planning was the Year 2 teacher's responsibility (47%) versus the TA's responsibility (47%). Data from interviews corroborated these findings: interviewees generally reported that the day-to-day delivery of Lexia fell to TAs, while ongoing monitoring was more often the responsibility of Year 2 teachers and/or headteachers.

The endline survey also asked headteachers to provide an open-text response about how well their school's division of tasks worked. The importance of having a cohesive team of dedicated staff delivering Lexia was a common theme in the responses received: 12 headteachers emphasised the benefits of having consistent, dedicated TAs or teachers delivering Lexia, while ten headteachers reported that their staff worked closely together to deliver Lexia. Several respondents also felt that involvement of the SLT was helpful: 11 headteachers noted that having a senior member of staff driving and monitoring the implementation of Lexia was useful. While most responses were positive, a minority of respondents reported challenges with the division of labour in their school. For instance, nine respondents noted that shortages and absence of staff trained to deliver Lexia was an issue.

Adaptations to Lexia delivery were usually not necessary, although some schools changed session timing

In their responses to the endline survey, a majority of headteachers (54%) and practitioners (68%) reported making no adjustments to their implementation of Lexia over the course of the trial. Among respondents who did report making adjustments, the most common adaptation was changing the time of day at which Lexia sessions were held. This was mentioned by eight headteachers and seven practitioners. The aim of this change was generally either to avoid pupils missing out on teaching time (in which case schools reported deciding to hold Lexia sessions during e.g. assembly or whole-class daily reading sessions), or to ensure that more pupils were able to attend (in which case schools moved Lexia sessions to later in the day). Another adaptation reported by several respondents was bringing in additional members of staff to deliver or oversee the implementation of Lexia, as noted by five headteachers and two practitioners.

Availability or reliability of technology could be a barrier, while having capable staff was a common facilitator

Barriers

Schools appear to have experienced few barriers to implementing Lexia. In the endline survey, headteachers and practitioners were asked which of a series of eight factors had been a barrier for their school. For every factor listed, a majority of both headteachers and practitioners stated that the factor in question had been 'not at all' a barrier.

That said, a large minority of headteachers and practitioners felt that insufficient or unreliable technology had, to some extent, made delivering Lexia harder. At endline, 46% of headteachers and 24% of practitioners agreed that insufficient hardware had been a barrier to a 'large', 'moderate', or 'small' extent, while around a third of both headteachers and

practitioners felt the same about unreliable internet connectivity. This was a point that also came up frequently in both interviews and open-text survey responses. For instance, in the endline survey, seven headteachers left open-text comments highlighting insufficient or poorly functioning devices or IT systems as a barrier, and four specifically mentioned having difficulties with the headphones that pupils wore when using Lexia. Difficulties with devices, particularly headphones, was also a common theme in interviews with practitioners. Another relatively frequently mentioned technology-related barrier in the endline survey was inadequate digital skills on the part of pupils, cited by 39% of headteachers and 27% of practitioners.

In addition to barriers associated with technology, some headteachers and practitioners highlighted the difficulty of finding time to deliver Lexia in the already packed school day. 14 headteachers left open-text responses to this effect in the endline survey. In interviews, several practitioners agreed that time was a barrier, but focused less on the difficulty of scheduling Lexia sessions than on the time investment required to set-up equipment ahead of sessions, fix any IT issues, and complete the additional paper-based exercises recommended by Lexia with pupils, which a number of practitioners found particularly time-consuming.

Finally, in interviews, some schools identified specific barriers for SEND pupils and learners who struggled with concentration. These respondents viewed Lexia as challenging for pupils who struggled to focus, especially when required to engage with the programme independently for 20 or 30 minutes at a time.

Facilitators

In open-text survey responses and interviews, schools highlighted a wide range of factors which had facilitated their delivery of Lexia. Capable school staff was the facilitator most often mentioned by headteachers: in the endline survey, 25 headteachers highlighted the benefits of having dedicated staff members delivering and monitoring Lexia, with nine emphasising the importance of having sufficient TAs and support staff in particular. Other facilitators noted by headteachers, and corroborated by practitioners in interviews, included: having access to suitable devices and headphones, mentioned by 15 headteachers; the high quality of training and support provided by Lexia, noted by 12 headteachers; setting dedicated time aside for delivering and/or monitoring Lexia, mentioned by nine headteachers; having a designated space for Lexia sessions, noted by six headteachers; and pupils' clear enjoyment of and engagement with Lexia, also mentioned by six headteachers.

Usual practice

This section discusses the following questions:

IPE RQ7: To what extent do school leadership, culture, and organisational structures influence the implementation of Lexia? What school-level policies and supports are in place to promote the use of digital technology in reading instruction and broader instructional practice?

IPE RQ8: What does business as usual look like when it comes to reading instruction for struggling readers?

IPE RQ9: How does Lexia differ from usual reading practice, which includes other targeted interventions that schools may implement as part of their business as usual?

Usual reading practice varied in observed control schools

Observations of Year 2 reading sessions revealed that a wide variety of session formats were utilised by control schools. These included whole-class reading sessions, small-group sessions, one-on-one sessions, and sessions in which pupils cycled between group work and individual work. Many schools integrated commercial interventions into their reading sessions. Session durations ranged from 20–30 minutes (for one-on-one and small-group sessions) to 45–50 minutes (for whole-class sessions). In every case, the session took place during regular class time. The most common location for observed sessions was the pupils' regular classroom. However, in several cases, sessions took place in other locations, such as the school library, the school hall, or a corridor. These settings were sometimes noisy or distracting for pupils.

Most schools were observed to provide some degree of differentiated instruction for readers at different levels. This was usually achieved by grouping pupils according to prior reading attainment, then assigning each group tasks appropriate to their level. Some schools provided one-on-one support, usually through TAs, for pupils who were particularly struggling with reading. Observers felt that the success of reading instruction seemed to rely on the skills and dedication of practitioners. In all cases, teachers and TAs were observed to be providing instruction, assistance, and support to pupils as needed, to the extent that was possible given teacher–pupil ratios.

Many schools considered their resources for struggling readers to be adequate

The majority of practitioners (85% at baseline) agreed or strongly agreed that their school provided adequate resources to enable them to give additional support to struggling readers. Free-text responses provided examples of this support and included having—or planning to introduce—reading interventions, schemes, or programmes. Other examples that were discussed by multiple respondents included support and resources, strategies for identifying and assessing struggling readers, capable staff, and staff training on teaching reading. Some noted that their schools used technology to support struggling readers, while others mentioned that they would like to use technology more for this purpose. Where respondents felt they could provide better or more support, cost and staff time were considered the largest barriers. Having an adequate number of TAs was also identified as an issue in some schools.

Many practitioners already used digital technology in their classrooms

Most schools used technology prior to participating in this trial. At baseline, 82% of headteachers reported that all of their teachers used digital technology. Only 2% of headteachers stated that just a few of their teachers used digital technology, and no headteachers stated that none of their teachers used it. At baseline, practitioners reported that the most common ways they used digital technology were delivering instruction (89%), planning lessons (82%) monitoring pupils' progress (77%), and providing additional support to struggling readers (66%). Just over half of practitioners (54% at baseline) reported accessing the digital technology they used in their classroom through a website rather than a mobile application.

Most schools had suitable technology

The majority of schools within both treatment and control groups seemed to have adequate technology in place at the start of the trial. Internet connections were considered suitable for supporting the use of digital technology. In baseline surveys, 71% of practitioners stated that their school's broadband connection was able to support their use of digital technology 'mostly, with rare connectivity issues'. A fifth responded that this was 'fully, with no connectivity issues' and no practitioners described this as 'inadequately, with frequent connectivity issues'. Similar proportions were reported in endline surveys. In interviews, issues with connectivity were noted by a handful of schools, but these were usually resolved with the help of the school's IT team.

Computers were available for most pupils, with 30% of practitioners reporting that their school had one device for each pupil, and 27% reporting that their school had one device per five pupils. However, 17% of practitioners reported having less than one device per 20 pupils.

Most practitioners agreed that their needs were 'often' or 'always' met by the technology in their classrooms, but this varied by activity. Practitioners were most likely to agree that technology met their needs when they were monitoring pupils' progress (75%), delivering instruction (65%), and planning lessons (62%). They were least likely to agree that technology met their needs when they were providing support to struggling readers: 46% of practitioners reported that the technology they used was 'always' or 'often' adequate for this purpose, with 35% choosing 'sometimes' and 11% 'rarely'.

The use of digital technology appeared to be well supported in the majority of treatment and control schools, with most schools having suitable technology in place prior to the start of the trial. The specific impact this is likely to have had on the implementation of Lexia is discussed below (see the 'Context, moderators, and mediators' section below). However, it is possible that schools that signed up to this trial were particularly interested in Educational Technology (EdTech), so may not be representative of all schools.

Cost restricted use of technology in some schools

The cost of digital technology appeared to be by far the most common factor preventing its use in schools: 82% of headteachers stated at baseline that the cost of digital technology was a barrier to its use in their school to a 'large' or 'moderate' extent. In interviews, participants highlighted the high cost of both new equipment and EdTech subscriptions.

When specifically considering reading, 45% of practitioners reported at baseline that they did not use digital technology to support reading instruction in the classroom. The reasons for this varied, but the most frequently selected responses were 'there are not enough devices available in my school' (22%), 'the necessary technology to use digital software is too expensive' (12%), and 'my students do not have sufficient digital skills to do so' (12%).

Interventions other than Lexia were quite common

In the year prior to the trial, around half of schools had implemented some sort of intervention or programme designed to improve Year 2 pupils' reading outcomes (reported by 51% of practitioners in the baseline survey). These interventions were described as typically used by 25 pupils or less (62%) or between 26 and 50 pupils (26%) per school and were delivered as small-group sessions (42%), one-to-one sessions (26%), or in a combination of two or more modes (15%). The majority of interventions were teacher-led (67%), with fewer being technology-led (20%) or a combination of both (13%). They were delivered before or after main classroom activities (40%), through a mixture of tasks completed both during main classroom activities and outside of them (36%), or during or alongside main classroom activities (24%). Practitioners reported that these interventions typically lasted for more than 24 weeks (55%), 4–8 weeks (16%), or 9–12 weeks (13%). Delivery times varied, between 31–60 mins per week (35%) and 30 mins or less (22%) being the most reported.

In free-text survey response items and interviews, headteachers and practitioners provided more information on specific interventions and programmes; however, some headteachers mentioned that their schools provided another form of systematic support for struggling readers, such as one-on-one reading sessions. Some headteachers described the ways in which their reading programme or intervention supported reading. Support with phonics was mentioned most frequently, followed by fluency and comprehension. A number of headteachers stated that their school's programme or intervention was bespoke, based on individual pupils' assessment results, and/or delivered one-on-one.

Most practitioners (82%) reported receiving training to deliver reading programmes or interventions. This training lasted less than one day for around two-thirds of teachers and TAs. Headteachers provided information on the number of members of staff who received training. In 40% of schools, fewer than five teachers had been trained to deliver the programme, and in 28% of schools, between five and nine teachers had been trained.

Compared to usual reading practice, Lexia was more technology-focused

Compared to usual practice, including other supplementary programmes and interventions, Lexia was focused on computer-based learning, typically within small groups, often outside of regular class time. This was seen in treatment schools, where all sessions observed were delivered in a small-group format. Observers did not systematically record the number of pupils in a session, but where this was noted, figures ranged from 4 to 16 pupils. Most sessions were between 20 and 30 minutes in duration, with some slightly longer. Overall, more time was typically spent on Lexia than on other interventions (typically at least 120 minutes a week). While the majority of sessions took place during regular class time, some treatment schools held Lexia sessions before the start of the school day or during the pupils' playtime. This perhaps reflects the fact that treatment schools tended to make Lexia a priority, putting protected time aside for Lexia sessions to ensure their pupils met the time targets. Unlike control group sessions, Lexia sessions were most often delivered in a location other than the pupils' usual classroom. Again, this perhaps reflects schools' commitment to the Lexia trial and willingness to assign the Lexia group a dedicated space.

Observations suggest that, at least in trial conditions, Lexia was delivered more formally and consistently than usual reading practice. As detailed above, treatment schools were more likely than control schools to allocate a dedicated time and space for pupils to complete Lexia sessions. Treatment schools were also more likely than control schools to have established procedures for pupils to follow at the start and end of each session.

Multiple practitioners agreed that overall, Lexia had been less staff-intensive than other interventions they had implemented. Some of them remarked that Lexia took longer to set-up than other interventions; however, once set-up, pupils mostly only needed supervision, and Lexia seemed ‘child-centred’ rather than ‘adult-centred’. Some respondents found that fitting-in Lexia sessions worked well because pupils were motivated; however, a similar number also anticipated that running Lexia would be more time-consuming with a larger number of pupils as guidance was still needed from staff, such as one-on-one lessons suggested by Lexia. Practitioners’ views on Lexia’s value for money varied considerably: several schools reported that Lexia would be unaffordable for them outside of the trial, while one respondent planned to rollout Lexia across a whole trust of schools. For some, Lexia was perceived as good value for money compared to previous interventions. Some respondents saw Lexia as individualised while others did not notice a significant difference, noting that reading classes were differentiated already. Although most control schools were observed to differentiate their reading instruction to some extent, only a minority were observed to provide one-to-one instruction to struggling readers. Lexia provides personalised instruction that is tailored to each individual pupil’s level, an advantage which was difficult for teacher-led interventions to replicate. Overall, there was a sense that Lexia was a very popular intervention among pupils, which made implementation easier.

In interpreting these findings, it should be noted that observers were only able to capture a snapshot of schools’ reading practice and may simply have been unaware of aspects of support being provided: for example, one-to-one instruction may have been taking place at a different time and in a different space. To some extent, many control sessions were therefore not directly comparable with Lexia sessions: for instance, whole-class interventions were not easy to compare to small-group Lexia sessions. It should also be noted that due to schools interpreting ‘supplementary programmes or interventions’ differently, for example, as whole-school or whole-class interventions, and due to the different nature of some of these interventions, direct comparisons with Lexia are difficult to make.

Perceived outcomes

This section discusses the following questions:

IPE RQ 11: In what ways, if any, do practitioners use Lexia progress reports to make decisions about additional support provided to struggling readers?

IPE RQ 12: How do practitioners scaffold instruction in practice and to what extent does Lexia influence this process?

IPE RQ 13: To what extent did the programme improve the confidence and ability of practitioners to support struggling readers?

Lexia was felt to be effective for struggling readers

Headteachers and practitioners generally felt that Lexia was very effective as an intervention for struggling readers. More than 80% of the headteachers and practitioners who responded to the endline survey agreed or strongly agreed that Lexia was effective at addressing the individual needs of struggling readers and that pupils’ reading ability had improved. At the same time, only around a tenth of headteachers and practitioners felt that using software at school had detracted from pupils’ learning (although around a third of practitioners neither agreed nor disagreed with this statement).

Most headteachers and practitioners also reported that their pupils had responded positively to Lexia and experienced wider benefits from use of the programme. Again, over 80% of endline survey respondents agreed or strongly agreed that pupils had engaged well with the programme and that pupils’ confidence in their reading had improved. Data from interviews painted a similar picture: the majority of interviewees reported that pupils engaged well with Lexia, were making good progress in reading, and in many cases had gained confidence in their reading ability through use of Lexia.

A small minority of survey respondents and interviewees noted caveats. These included: waning enthusiasm for Lexia among pupils as the year went on; a sense that Lexia worked better for some pupils than others, and that certain pupils, particularly those with additional needs or EAL, struggled to engage with the programme; and doubts about the impact of Lexia on pupils’ reading ability. However, it is worth emphasising that each of these caveats was mentioned by only a handful of respondents, and that feedback on the programme was in general extremely positive.

Practitioners were already confident teachers of reading, but less experienced staff saw confidence gains

The potential impact of Lexia on practitioners' confidence in supporting pupils to acquire reading skills was a key short-term outcome in the programme's theory of change. To some extent, the hypothesis that Lexia would build practitioners' confidence is borne out by our findings: as this section discusses, our survey data suggests that practitioners' confidence did indeed increase over the period of the trial.

However, the results of the baseline survey demonstrate that practitioners were in fact already confident about teaching reading when the trial began. At baseline, practitioners were asked about their confidence in delivering each of a series of 13 instructional activities, including administering and using the results of reading assessments, scaffolding reading instruction, and strengthening pupils' reading ability and their interest and confidence in reading. The vast majority of practitioners responded that they were 'confident' or 'very confident' about delivering each of the 13 activities.

Endline survey responses suggest that, while both TAs and Year 2 teachers were already confident teachers of reading, TAs saw a larger boost to their confidence from delivering Lexia than Year 2 teachers. Both practitioners and headteachers were asked about changes in practitioner confidence at endline. 63% of practitioners felt that the designated TA had shown increased confidence in teaching struggling readers over the course of the trial. However, only 39% of practitioners noted an increase in the confidence of the Year 2 teacher, with 44% neither agreeing nor disagreeing that the Year 2 teacher's confidence had increased. Data collected from headteachers showed a similar pattern: around 60% of headteachers had seen an increase in the designated TA's confidence in teaching struggling readers, but only 42% of headteachers had noted a similar change in their Year 2 teacher, with over a third neither agreeing nor disagreeing that the Year 2 teacher's confidence had improved.

Finally, practitioners were also asked about their personal level of confidence with the use of digital software in teaching reading. At endline, just over half of respondents stated that their confidence level had improved since they began implementing Lexia. This suggests that many practitioners were already confident users of EdTech when the trial began. Indeed, several practitioners made comments to this effect when asked to elaborate on their survey response, and a number of interviewees made the same point.

Teachers had mixed views on whether Lexia had improved their school's teaching of reading

In both surveys and interviews, headteachers and practitioners were asked about the impact of Lexia on the way reading was taught in their school and on practitioners' ability to provide support to struggling readers. Their responses were mixed. When asked in the endline survey whether they had seen positive changes in the teaching of reading in their school since implementing Lexia, 44% of headteachers agreed, 32% neither agreed nor disagreed, and 20% disagreed. Practitioners' views were equally diverse: 32% agreed that the teaching of reading in their class had improved; 40% neither agreed nor disagreed; and 19% disagreed.

One explanation for this diversity of opinions is that many schools were already implementing best practice in reading instruction. For this reason, Lexia may not have been perceived to meaningfully improve teaching practices. However, many practitioners nonetheless appear to have found Lexia helpful in facilitating these practices. For example, at baseline, more than 90% of practitioners already felt 'confident' or 'very confident' in their ability to scaffold reading instruction based on their pupils' skill level. Nevertheless, at endline, 79% of practitioners reported that Lexia had, to either a 'small', 'moderate', or 'large' extent, helped them with scaffolding. Similarly, more than 90% of practitioners reported at baseline that they used the results of formative assessments to adjust their instruction to each pupil's current level of skill. Nonetheless, 67% of practitioners and 78% of headteachers reported using the progress reports generated by Lexia to adjust instruction to pupils' individual needs. In short, while Lexia may not have drastically changed treatment schools' reading practices, staff nonetheless appear to have used the programme to support these practices.

Staff members' views on the impact of Lexia on their workload and ability to spend time supporting pupils one-on-one, were similarly equivocal. At endline, more than three-quarters of headteachers in treatment schools felt that using software had allowed practitioners to provide personalised support to more pupils than before. There was also a small increase from baseline to endline in the percentage of practitioners (across both control and treatment schools) who agreed that they were able to provide supplementary one-to-one instructional support to pupils who struggled with reading (68% at baseline

vs 73% at endline). However, this change does not appear to have come about because Lexia reduced practitioners' workload: only a minority (18%) of practitioners in treatment schools agreed that delivering Lexia had freed up their time for more meaningful instructional activities; while 39% neither agreed nor disagreed; and 32% disagreed. At the same time, Lexia does not seem to have created an additional burden: two-thirds of practitioners disagreed that delivering Lexia had significantly enlarged their workload.

Context, moderators, and mediators

This section discusses the following questions:

IPE RQ2: In practice, which contexts and mechanisms mediate the intervention logic's causal pathways?

IPE RQ7: To what extent do school leadership, culture, and organisational structures influence the implementation of Lexia? What school-level policies and supports are in place to promote the use of digital technology in reading instruction and broader instructional practice?

Lexia implementation was supported by commitment to EdTech

Support for digital technology was an important context for schools and enabled successful implementation. As discussed here above (in the 'Usual practice' subsection), the use of digital technology appeared to be well supported, with the majority of schools having appropriate technology in place.

Among Lexia schools, interviews and observations revealed that school culture and policies influenced the implementation of Lexia in various ways. Commitment from school leadership was mentioned as important in supporting the adoption of EdTech interventions. Examples of leadership support for Lexia included sending letters to parents/legal guardians of participating pupils, completing costing exercises, and celebrating Lexia successes in assemblies by handing out certificates to pupils. This aligns with Lexia's theory of change in terms of buy-in from senior management to support practitioners.

Although many schools reported what appeared to be relatively high ratios of pupils to devices (as discussed here in the 'Usual practice' subsection), interviews with Lexia practitioners revealed that a lack of devices, or difficulty acquiring devices for Lexia sessions, could sometimes be problematic. This was the most frequently cited obstacle, with several schools reporting having to book Chromebooks or iPads in advance. This highlights the importance of ensuring that schools implementing an EdTech intervention have appropriate technology available as needed.

Having dedicated time and space for Lexia sessions supported implementation

Implementation of Lexia was aided by staff delivering sessions more formally and consistently than usual reading practice. Organising dedicated spaces for Lexia sessions was important in ensuring delivery could run smoothly and in a suitable setting. In comparison to some observed reading sessions in control schools, where locations such as corridors were used and found to be noisy and distracting for pupils, Lexia sessions tended to be run in a more suitable setting. Treatment schools were also more likely than control schools to set aside dedicated time for Lexia sessions and have established procedures for pupils to follow at the start and end of each session.

Training helped support practitioners, especially when they were less experienced

Although Lexia is computer-led, observations and interviews with headteachers highlighted the importance of having experienced and dedicated staff. Practitioners were able to support pupils and ensure sessions ran efficiently by completing tasks such as setting out equipment before sessions began, managing behaviour, helping with re-focusing any distracted pupils, and supporting with technology where needed.

As discussed in interviews, some schools had little experience with other reading interventions and EdTech, and some had only recently purchased equipment. A related issue was the nervousness of staff around technology, as reported by some schools. In line with Lexia's theory of change, staff training was well attended and supported staff members' understanding of the software and confidence in instructing pupils. Lexia's training and support were valued by headteachers and

practitioners, as reflected in surveys and interviews. This—combined with involvement in Lexia—appeared to contribute to building practitioners’ confidence in delivering EdTech interventions, particularly for TAs and those with less experience.

Lexia’s personalised, technology-led instruction was important in supporting struggling readers

The personalised, technology-led nature of Lexia instruction appeared to differentiate this intervention from others, including usual classroom instruction. Data from interviews and observations suggest that most schools were already differentiating their reading instruction to some extent, including providing at least some one-on-one support to pupils who particularly needed it (e.g. pupils with SEND, pupils attracting the Pupil Premium grant, or pupils who were particularly struggling with reading). However, particularly for schools with fewer TAs, this support could be minimal and/or difficult to provide as regularly as schools would have liked. Lexia enabled schools to provide tailored instruction that was appropriate to each individual pupil’s level both more regularly and to more pupils at once. This degree and extent of personalisation was difficult to replicate in teacher-led interventions.

The digital and personalised nature of Lexia also appeared to support educators in various ways. For example, multiple schools noted that data from Lexia was valuable in saving teachers time spent on tasks such as marking and assessment through the use of Lexia progress reports. In line with the theory of change, by using Lexia, pupils were able to spend more time practising elements that supported the acquisition of reading skills at a level appropriate for their level of development. The outcome of this was an improvement in reading in Lexia schools compared to control schools, as identified in the impact evaluation.

Cost

The evaluation team collected cost data from LexiaUK and from participating schools (via surveys and interviews) with the intention of establishing the average costs of implementation of Lexia Reading Core5. We estimate average costs per pupil per year for schools implementing the Lexia programme based on the maximum licence usage but have also noted costs specifically for pupils involved in this trial.

The main financial costs of implementation were:

- the Lexia three-year licence package; and
- any additional costs for facilities, equipment, or materials.

Additional time costs considered were for:

- staff training; and
- practitioner delivery time.

Direct costs

The total cost of a three-year Lexia entry package was £3,850 (excluding VAT, which schools can claim back, so it is not included in these costings) (Table 29). This included 50 licences; however, licences could be reallocated over the licence period, so considerably more than 50 pupils could use Lexia over the three years. For example, original pupils could be removed if they reached age-related expectations and new pupils could be added.

The licence package also included four initial one-hour training sessions and one annual training session for each year for the remainder of their licence (i.e. six training sessions over the three-year licence period). In addition, there was unlimited support available by phone or Microsoft Teams. Unlimited extra training and support beyond the six sessions could be requested by schools at any time at no additional cost. Hence there were no additional direct costs for training.

Following the collection of cost data from Lexia, we have based three-year costs on the assumption that the 50 pupils using licences would change each year (although in theory they could change more frequently than this). Therefore, three-year costs of £25.67 per pupil (Table 29) have been estimated, based on 50 different pupils using Lexia in each of the three years (i.e. 150 pupils in total).

Within this trial, the average (mean) number of pupils who used Lexia was 10.2. However, in surveys and interviews, schools noted that more pupils were added outside of the trial, to maximise their licences. As those numbers were not recorded by Lexia or shared by schools, a cost of £377.45 for each pupil involved in the trial, per school, was estimated. However, it should be noted that this is likely a considerable underestimate of the total number of pupils using Lexia within trial schools, hence an overestimate of costs per pupil.

Additional costs

The majority of headteachers (61%) stated that their school had not incurred any additional costs associated with procuring facilities, equipment, or materials required to deliver Lexia. Given that delivery is online, the need for printed resources was minimal, but some schools did need to purchase equipment such as tablets or headphones for the trial. Of those who incurred costs, 61% of headteachers reported the total cost of procuring additional facilities, equipment, and materials required to deliver Lexia to be £250 or less. However, a smaller number of headteachers reported larger costs, with 13% spending over £1,250. Overall, additional costs associated with Lexia are estimated at an average of £335.63 (Table 29), although it should be noted that the median value is £0.¹⁰ Given these costs were typically for equipment that was purchased at the start of the trial and would not need to be replaced during the trial, these costs have been included once (i.e. they are not repeated in years two or three). Although delivery of Lexia to greater numbers of pupils might mean some increase in the amount of equipment needed, given the small-group nature of Lexia and the rotation of equipment observed

¹⁰ The mean was affected by one very high outlier.

in schools, these costs are considered minimal and have not been estimated. It is also important to note that this equipment could also be used for other purposes, so in that sense, any estimated costings would likely be higher than would actually be the case for most schools.

Time costs

Across all treatment schools, a total of 353 educators took part in training. On average, three staff members took part in training per school. These were typically teachers, TAs, and in most cases, headteachers and/or senior leaders, who attended to understand the programme and support its implementation. There were four online training sessions per school, spread over the course of the study period. The first two sessions were often delivered on the same day. Each training session lasts for approximately one hour. Data were not collected on whether cover was needed for staff attending training, but it would seem sensible to assume that this was potentially needed for teachers and TAs. If this involved two or three practitioners per school attending approximately four hours of training, an estimate of 8 to 12 hours of cover may be required per school. Although not included within the trial period, there would potentially be an additional two to three hours of training in each subsequent year per school (an additional hour's training each year for each of the two or three staff members involved in Lexia implementation). In total, training time is estimated at 12 to 18 hours per school, depending on the number of staff attending. Costs would depend on whether cover was needed for members of staff attending training.

Most practitioners (92%) reported that they spent 150 minutes or less in contact with pupils to deliver Lexia each week. Roughly equal numbers of headteachers reported that implementing Lexia had (53%) and had not (47%) required teacher cover for staff members. Of those headteachers who stated that teacher cover was required related to the implementation of Lexia, just over half (52%) reported this to be £250 or less, 29% reported this to be between £251 and £500, and the remaining 19% reported spending over £750. Overall, cover costs related to implementing Lexia are estimated at £315.14 (mean; £90 median).¹¹ These time estimates are based on trial pupil numbers (on average ten per school) and it should be noted that staff delivery time would likely increase as the number of pupils involved in Lexia grew. We do not have data to support estimates of this and, in keeping with the EEF cost guidance (EEF, 2023), we have not included estimates of these costs within our overall cost estimates.

Table 29: Cost of delivering Lexia

Item	Type of cost	Cost	Total cost over three years	Total cost per pupil per year over three years
Three-year licence fee	Start-up cost per school (excluding VAT, which schools can claim back)	£3,850	£3,850	£25.67
Facilities, equipment, and materials	Set-up and/or running cost per school	£335.63	£335.63	£2.24
Total			£4,185.63	£27.91

Table 30: Cumulative costs of Lexia (assuming delivery over three years)

	Year one	Year two	Year three
Lexia delivery at maximum licence usage per school	£4,185.63	£4,185.63	£4,185.63

¹¹ As above, this estimate was affected by high outliers.

Cost of business as usual in control schools

The cost of business as usual reading interventions for struggling readers in Year 2 was estimated by analysing survey responses provided by headteachers in control schools on the cost of the supplementary reading programme or intervention for Year 2 pupils most recently implemented in their school. Where programmes had been implemented, in almost half of schools (48%), headteachers reported the total direct cost of the programme to be less than £1,000. In 14% of cases, the total cost was more than £5,000. In most cases (81%), the programme required cover for staff members involved in delivering it. The cost of providing this cover was less than £1,000 in the majority (54%) of cases. 72% of headteachers reported no additional costs related to implementing other interventions. Where there were additional costs, a quarter of headteachers stated that these came to less than £1,000, and another quarter reported costs of between £1,000 and £1,999. In the remaining 50% of cases, additional costs came to £2,000 or more.

For several reasons, caution should be exercised in interpreting this data. First, the figures given above are self-reported and have not been validated by the research team. Second, headteachers interpreted the term ‘supplementary reading programme or intervention for Year 2 pupils’ differently and therefore, provided data on programmes with varying scope and reach. Direct comparisons between the cost of a Lexia licence and the cost of business as usual as described above are therefore unlikely to be robust.

Conclusion

Table 31: Key conclusions

Key conclusions	
1.	Pupils in Lexia schools made two additional months' progress in reading, on average, compared to pupils in other schools. This result has a high security rating.
2.	Among pupils eligible for FSM, those in Lexia schools made two additional months' progress in reading, on average, compared to those in control schools. These results may have lower security than the overall findings because of the smaller number of pupils.
3.	Pupils in Lexia schools made two additional months' progress in the following specific reading skills: Word Identification, Applying Phonics, and Passage Comprehension, on average, compared to pupils in control schools.
4.	Exploratory evidence suggests that increasing usage of Lexia beyond the minimum guidelines leads to greater progress in reading proficiency.
5.	Lexia was relatively easy to implement in schools which had the appropriate technology available, e.g. laptops, headphones, and internet.

Impact evaluation and IPE integration

The results of the evaluation provide strong support for Lexia's theory of change

Taken together, the results of the impact evaluation and IPE provide strong support for most elements of Lexia's theory of change, including the model's inputs, outputs, and pupil- and teacher-level outcomes.

The IPE findings and Lexia usage data suggest that most schools completed all of the inputs and outputs included in the theory of change. Most struggling readers used Lexia as outlined in the theory of change, spending at least the minimum amount of time required on the programme (and frequently much more) and working individually at a level appropriate for them. Practitioners attended training, facilitated Lexia sessions as intended, and provided additional one-on-one support to struggling readers where necessary, while both practitioners and headteachers reported using the data generated by Lexia on pupils' progress.

Evidence for pupil-level outcomes is particularly strong. The impact evaluation found that struggling readers in treatment schools made two months' relative progress compared to struggling readers in control schools, although it is worth noting that age-adjusted scores fell in both the treatment and control groups, suggesting that the pupils who were assessed still made less than expected progress. This is strong evidence for the effectiveness of Lexia in supporting pupils' reading proficiency. Most headteachers and practitioners too perceived Lexia to be an effective intervention for struggling readers: more than four-fifths of endline survey respondents felt that pupils had engaged well with Lexia and seen improvements in their reading.

This evaluation has also generated evidence for the theory of change's teacher-level outcomes. One of the theory of change's short-term outcomes was that the intervention would generate buy-in from schools' SLTs. Findings from interviews and endline surveys support this: most headteachers were enthusiastic about the benefits of Lexia, and in interviews, practitioners often reported that their school's senior leaders had supported them to implement Lexia. The evidence for practitioner-level outcomes is perhaps less strong. The theory of change hypothesised that delivering Lexia would improve practitioners' confidence in supporting struggling readers with their reading. To some extent, this appears to have been the case: a majority of endline survey respondents felt that the TA responsible for delivering Lexia sessions had grown in confidence over the course of the trial. However, only a minority of respondents felt the same about the Year 2 teacher, and baseline survey results suggested that the vast majority of practitioners involved in Lexia (both TAs and teachers) were already confident in their ability to teach reading before the trial began.

Lexia is effective for Year 2 pupils, straightforward to deliver, and may be suitable for scale-up

This evaluation has found strong evidence that Lexia is beneficial for the reading outcomes of struggling readers in Year 2. As discussed above, while neither the control nor the treatment group appear to have made as much progress as expected in reading, struggling readers in treatment schools made more relative progress compared to struggling readers in control schools. This result is a robust contribution to the considerable existing evidence for the effectiveness of Lexia (Lexia®, 2024), which mostly relates to use of Lexia in the US. It also adds to the more limited evidence base for the use of computer-assisted instruction to improve pupils' language skills in UK schools (see e.g. O'Callaghan *et al.*, 2016).

The evaluation also builds on the EEF previous efficacy trial of Lexia (Tracey *et al.*, 2022). The efficacy trial found a significant difference in reading outcomes between pupils who received Lexia and those who did not, although with a modest effect size of 0.08, equating to one month's additional progress. This evaluation also found a significant difference in reading outcomes between pupils in treatment and control schools, but with a greater effect size of 0.145. This evaluation therefore provides a higher estimate for the impact of Lexia on the reading ability of struggling readers in Year 2 than the previous efficacy trial.

The positive impact of Lexia was sustained for struggling readers who were classed as FSM and EAL, for whom the estimated effect size is equivalent to two months' and three months' relative progress, respectively. There is, however, insufficient evidence to suggest that Lexia is a gap-closer. Furthermore, additional evidence suggests that SEND pupils identified as struggling readers do not make statistically significant progress when using Lexia, and that struggling readers with higher baseline attainment benefit from the programme more than their lower baseline attainment. This suggests that in some respects, Lexia could be a gap-widener.

In many other respects, this evaluation produced similar findings to the efficacy trial. Compliance was high in both trials: 93% of participating pupils in the efficacy trial and 97% of pupils in the effectiveness trial spent at least the minimum required number of minutes on Lexia over the course of the trial. Indeed, in this trial, pupils were found to spend an average of 2,687 minutes using Lexia, more than three times the minimum required number of minutes. Implementation fidelity was also high in both trials. Taken together, the positive findings on compliance and fidelity from both trials suggest that Lexia is relatively straightforward for schools to deliver and incorporate into their existing routines, particularly where schools had sufficient technology available as needed.

The encouraging results of this evaluation suggest that scaling up schools' use of Lexia could benefit struggling readers. Indeed, given that exploratory analysis found larger impacts from use of Lexia for struggling readers with relatively higher baseline reading attainment, scale-up could also prove beneficial for the reading ability of pupils who are not struggling. However, in interviews, several practitioners and headteachers noted potential barriers to scale-up. These included the cost of Lexia and the need for sufficient devices to enable many pupils to use Lexia at once. It is also possible that the findings of this trial may not fully generalise to all UK schools, as the schools who enrolled in this trial may have been more committed than average to the use of EdTech and had more robust digital infrastructure in place, which would in turn have facilitated their delivery of Lexia. Finally, as suggested by the exploratory analysis undertaken as part of this trial, by interviews with practitioners, and also by IPE evidence from the efficacy trial, Lexia may not be suitable for all pupils: those who struggle to focus or who have very low prior reading attainment may not benefit as much from Lexia or may need one-to-one support to use the Lexia programme effectively. These factors should be borne in mind if scale-up is being considered.

Limitations and lessons learned

The evaluation was subject to some minor forms of bias, but no major threats to validity

The research team are not aware of any major threats to either the internal or the external validity of the impact evaluation. That said, there is some limited evidence for a potential Hawthorne effect among treatment schools. The IPE found that, on average, treatment schools appeared to be delivering reading sessions (i.e. Lexia) more formally and consistently than control groups were delivering their business as usual reading sessions. It is therefore, possible that some of the observed effect of Lexia could be due to pupils in the treatment group receiving more regular and/or more highly prioritised reading instruction than pupils in the control group, and not linked to Lexia per se. However, it must be noted that this IPE finding is

based on a small number of observations of Lexia sessions and control group reading sessions, and that Hawthorne effects are a strong possibility in any non-blinded study, including the vast majority of education trials.

Another minor but noteworthy threat to the validity of the impact evaluation is the possibility of selection bias in the schools that signed up to participate in the trial. As noted above, it is possible that these schools may have been more likely than the average school to be enthusiastic about the use of EdTech, to have strong digital infrastructure already in place, and to have SLTs that were supportive of EdTech. For this reason, the schools who participated in the trial may have found Lexia more straightforward to implement than non-participating schools would have done. If true, this would limit the generalisability of the evaluation's results.

Access to detailed usage data could be helpful in evaluating future technology-led interventions

In future evaluations of technology-led interventions, it would be helpful for the research team to have access to more granular data on pupils' usage of the intervention, rather than simply the total minutes spent on the programme per pupil. For instance, in the case of this study, pupil-level data on the number and duration of individual sessions would have enabled a clearer picture to be built up of pupils' usage of Lexia. Access to this data would also have enabled more light to be shed on some of the study's findings. For instance, exploratory analysis conducted as part of this study suggested that Lexia may have a larger impact on pupils with higher reading proficiency at baseline, despite the fact that pupils with higher reading proficiency were found to spend similar amounts of time on Lexia to readers with lower baseline proficiency. Having access to data on pupils' engagement with Lexia could have enabled the research team to better understand this finding. For example, it is possible that more able readers spent more time actively engaging with and attending to the programme's activities than less able readers. This hypothesis could have been tested if activity counts (e.g. clicks/taps per session) were available.

Future research and publications

Further research is needed to understand the differential and longitudinal effects of Lexia

This study has contributed to an increasing body of evidence suggesting that Lexia can improve pupils' reading outcomes. However, many questions about the effectiveness of Lexia remain. For instance, while there is now strong evidence for the effect of Lexia on the reading attainment of struggling readers in Year 2, there is less evidence for the impact of Lexia on other populations, particularly in the UK context. For example, further research is needed to establish Lexia's effects on different year groups and on pupils of different ability levels. The latter question is particularly interesting in the light of this study's exploratory analysis of the effect of Lexia on Year 2 pupils with lower versus higher baseline reading attainment, which found that Lexia in fact appeared more effective for stronger, struggling readers.

The dosage-related effects found in both the efficacy trial and the effectiveness trial could also be useful to investigate further. For instance, exploratory analysis in this trial found that pupils who used Lexia more, experienced more benefit, with pupils who received five times the recommended dosage of Lexia (3,600 minutes vs 720 minutes) experiencing a full ten months' additional progress in reading (although, as noted above, this result may reflect selection bias, whereby pupils who benefit more from Lexia spend more time on the programme). Schools considering using Lexia more intensively would benefit from knowing whether there is a point of diminishing returns from Lexia use, at which pupils begin to experience less improvement in their reading from use of the programme.

Further research is required to establish whether the short-term progress that pupils have been shown to make by using Lexia is maintained in the longer term. A positive finding would confirm the value of Lexia for improving pupils' reading proficiency.

Finally, Lexia consists of a three-part model, which is a change from how the theory of change was initially operationalised. This evaluation has focused on independent, pupil-driven learning through the Lexia software activities. There are other dimensions to Lexia which need to be explored in future rollout, namely: the use of data, including reports, to monitor progress and identify pupils who may benefit from additional support; and the use and impact of further support, such as printable teaching resources designed to help accelerate reading skills in lower performing pupils.

References

- Angrist, J. (2006) 'Instrumental Variables Methods in Experimental Criminological Research: What, Why and How'. *Journal of Experimental Criminology*, 2: 1, 23–44. <https://doi.org/10.1007/s11292-005-5126-x>
- Angrist, J.D. and Krueger, A.B. (1991) 'Does Compulsory School Attendance Affect Schooling and Earnings?' *The Quarterly Journal of Economics*, 106: 4, 979–1014. <https://doi.org/10.2307/2937954>
- Babayigit, S. (2014) 'The Role of Oral Language Skills in Reading and Listening Comprehension of Text: A Comparison of Monolingual (L1) and Bilingual (L2) Speakers of English Language'. *Journal of Research in Reading*, 37: Suppl 1, S22–S47. <https://doi.org/10.1111/j.1467-9817.2012.01538.x>
- Bates, D., Maechler, M., Bolker, B. and Walker, S. (2015) 'Fitting Linear Mixed-Effects Models Using lme4'. *Journal of Statistical Software*, 67: 1, 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Braun, V. and Clarke, V. (2006) 'Using Thematic Analysis in Psychology'. *Qualitative Research in Psychology*, 3: 2, 77–101. doi: 10.1191/1478088706qp063oa
- Burgoyne, K., Whiteley, H. and Hutchinson, J. (2011) 'The Development of Comprehension and Reading-Related Skills in Children Learning English as an Additional Language and their Monolingual, English-Speaking Peers'. *British Journal of Educational Psychology*, 81: 2, 344–354. <https://doi.org/10.1348/000709910X504122>
- Carpenter, J., Kenward, M. and White, I. (2007) 'Sensitivity Analysis After Multiple Imputation Undermissing at Random: A Weighting Approach'. *Statistical Methods in Medical Research*, 16: 259–275. <https://doi.org/10.1177/0962280206075303>
- Carril, A. (2017) 'Dealing With Misfits in Random Treatment Assignment'. *The Stata Journal*, 17: 3, 652–667. <https://doi.org/10.1177/1536867X1701700307>
- Cheung, A.C.K. and Slavin, R. (2012) 'How Features of Educational Technology Applications Affect Student Reading Outcomes: A Meta-Analysis'. *Educational Research Review*, 7: 3, 198–215. doi: 10.1016/j.edurev.2012.05.002
- Clarke, D., Joseph Romano, J. and Wolf, M. (2019) *The Romano-Wolf Multiple Hypothesis Correction in Stata*. Bonn: IZA Institute of Labor Economics.
- Coe, R. (2002) *It's the Effect Size, Stupid. What Effect Size Is and Why It Is Important*. Paper presented at the Annual Conference of the British Educational Research Association Annual Conference, University of Exeter, 12–14 September 2002. Available at: https://internet.psych.wisc.edu/wp-content/uploads/532-Master/532-UnitPages/Unit-03/Coe_EffectSize_2002.pdf (accessed 23 01 2025).
- Data Protection Act 2018, c.12. Available at: www.legislation.gov.uk/ukpga/2018/12/contents (accessed 13 09 2026).
- Daniel, J. (2024) 'The Academic Achievement Gap Between Students With and Without Special Educational Needs and Disabilities'. *European Journal of Special Needs Education*, 40: 3, 539–556. doi: 10.1080/08856257.2024.2400771
- Dong, N. and Maynard, R. (2013) 'PowerUp!: A Tool for Calculating Minimum Detectable Effect Sizes and Minimum Required Sample Sizes for Experimental and Quasi-Experimental Design Studies'. *Journal of Research on Educational Effectiveness*, 6: 1, 24–67. <https://doi.org/10.1080/19345747.2012.673143>
- Education Endowment Foundation (EEF). (2021) *Teaching and Learning/Early Years Toolkit Guide*. London: Education Endowment Foundation. Available at: <https://d2tic4wvo1iusb.cloudfront.net/documents/toolkit/EEF-Toolkit-guide.pdf> (accessed 13 09 2026).
- Education Endowment Foundation (EEF). (2022a) *Implementation and Process Evaluation Guidance for EEF Evaluations*. London: Education Endowment Foundation. Available at:

<https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/evaluation-design/EEF-IPE-Guidance-August-2022.pdf?v=1751983127> (accessed 29 05 2026).

Education Endowment Foundation (EEF). (2022b) '*Statistical Analysis Guidance for EEF Evaluations*'. London: Education Endowment Foundation. Available at: <https://d2tic4wvo1iusb.cloudfront.net/documents/evaluation/evaluation-design/EEF-Analysis-Guidance-Website-Version-2022.14.11.pdf?v=1681842352> (accessed 29 05 2026).

Education Endowment Foundation (EEF). (2023) '*Cost Evaluation Guidance for EEF Evaluations*'. London: Education Endowment Foundation. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/evaluation-design/Cost-Evaluation-Guidance-Feb_2023.pdf?v=1751983127 (accessed 29 05 2026).

Ferrer, E., Shaywitz, B., Holahan, J., Marchione, K., Michaels, R. and Shaywitz, S. (2015) 'Achievement Gap in Reading is Present as Early as First Grade and Persists Through Adolescence'. *The Journal of Pediatrics*, 167: 5, 1121–1125.e2. <https://doi.org/10.1016/j.jpeds.2015.07.045>

Ge, X. and Land, S. (2003) 'Scaffolding Students' Problem-Solving Processes in an Ill-Structured Task Using Question Prompts and Peer Interactions'. *Educational Technology Research and Development*, 51, 21–38. <https://doi.org/10.1007/BF02504515>

General Data Protection Regulation (GDPR). (2016) '*Council Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data (United Kingdom General Data Protection Regulation) (Text with EEA relevance)*'. Available at: www.legislation.gov.uk/eur/2016/679 (accessed 13 09 2026).

Griffiths, Y. and Stuart, M. (2013) 'Reviewing Evidence-Based Practice for Pupils With Dyslexia and Literacy Difficulties'. *Journal of Research in Reading*, 36: 1, 96–116. <https://doi.org/10.1111/j.1467-9817.2011.01495.x>

Hartas, D. (2011) 'Families' Social Backgrounds Matter: Socio-Economic Factors, Home Learning and Young Children's Language, Literacy and Social Outcomes'. *British Educational Research Journal*, 37: 6, 893–914. <https://doi.org/10.1080/01411926.2010.506945>

Hedges, L. (2007) 'Effect Sizes in Cluster-Randomized Designs'. *Journal of Educational and Behavioral Statistics*, 32: 4, 341–370. doi: 10.3102/1076998606298043

Henk, W.A., Moore, J.C., Marinak, B.A. and Tomasetti, B.W. (2000) 'A Reading Lesson Observation Framework for Elementary Teachers, Principals, and Literacy Supervisors'. *The Reading Teacher*, 53: 5, 358–369. www.jstor.org/stable/20204805

HM Government. (2022) '*Levelling Up: Levelling Up the United Kingdom*'. GOV.UK. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1052706/Levelling_Up_WP_HRES.pdf (accessed 01 06 2026).

Hoge, R.D. and Coladarci, T. (1989) 'Teacher-Based Judgements of Academic Achievement: A Review of the Literature'. *Review of Educational Research*, 59: 3, 297–313. <https://doi.org/10.3102/00346543059003297>

Hutchison, D. and Styles, B. (2010) '*A Guide to Running Randomised Controlled Trials for Educational Researchers*'. Slough: NFER. Available at: www.nfer.ac.uk/media/qdjkwn3i/rct01.pdf (accessed 03 11 2025).

Jakobsen, J.C., Gluud, C., Wetterslev, J. and Winkel, P. (2017) 'When and How Should Multiple Imputation be Used for Handling Missing Data in Randomised Clinical Trials – A Practical Guide With Flowcharts'. *BMC Medical Research Methodology*, 17: 1, 1–10. <https://doi.org/10.1186/s12874-017-0442-1>

Kulik, J. (2003) '*Effects of Using Instructional Technology in Elementary and Secondary Schools: What Controlled Evaluation Studies Say*'. Arlington, VA: SRI International.

- Lenkeita, J., Knut Schwippert, K. and Knigge, M. (2018) 'Configurations of Multiple Disparities in Reading Performance: Longitudinal Observations Across France, Germany, Sweden and the United Kingdom'. *Assessment in Education: Principles, Policy & Practice*, 25: 1, 52–86. <https://doi.org/10.1080/0969594X.2017.1309352>
- Lewin, C., Smith, A., Morris, S. and Craig, E. (2019) '*Using Digital Technology to Improve Learning: Evidence Review*'. London: Education Endowment Foundation. Available at: <https://educationendowmentfoundation.org.uk/education-evidence/guidance-reports/digital> (accessed 13 09 2026).
- Lexia®. (2024) '*Product Evidence Base: Lexia® Core5® Reading Efficacy Research*'. Available at: www.lexialearning.com/user_area/content_media/processed/Core5EfficacyResearch1.pdf (accessed 20 10 2025).
- Lüdecke, D., Ben-Shachar, M.S., Patil, I., Waggoner, P. and Makowski, D. (2021) 'Performance: An R Package for Assessment, Comparison and Testing of Statistical Models'. *Journal of Open Source Software*, 6: 60, 3139. <https://doi.org/10.21105/joss.03139>
- Merewood, J., Senator, B., Hutton, E., Reed, M., Subosa, M. and Speciani, E.R. (2025) '*Independent Evaluation of the Lexia® Core5® Reading Programme: A Cluster Randomised Controlled Trial: Evaluation Protocol*'. Available at: <https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/lexia-core5-reading-2022-2023-trial> (accessed 20 10 2025).
- Northrop, L. (2017) 'Breaking the Cycle: Cumulative Disadvantage in Literacy'. *Reading Research Quarterly*, 52: 4, 391–396. <https://doi.org/10.1002/rrq.195>
- O'Callaghan, P., McIvor, A., McVeigh, C. and Rushe, T. (2016) 'A Randomized 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, 546–558. <https://doi.org/10.1111/bjep.12122>
- Pawson, R. and Tilley, N. (1997) 'An Introduction to Scientific Realist Evaluation'. In: E. Chelimsky and W.R. Shadish (Eds.), '*Evaluation for the 21st Century: A Handbook (pp. 1–150)*'. Thousand Oaks, CA: SAGE Publications, Inc.
- R Core Team. (2025) '*R: A Language and Environment for Statistical Computing*'. Vienna: R Foundation for Statistical Computing. Available at: <https://cran.r-project.org/doc/manuals/r-release/fullrefman.pdf> (accessed 13 09 2026).
- Russell, G., Ukoumunne, O., Ryder, D., Golding, J. and Norwich, B. (2018) 'Predictors of Word-Reading Ability in 7-Year-Olds: Analysis of Data from a UK Cohort Study'. *Journal of Research in Reading*, 41: 1, 58–78. <https://doi.org/10.1111/1467-9817.12087>
- Scammacca, N., Fall, A.-M., Capin, P., Roberts, G. and Swanson, E. (2020) 'Examining Factors Affecting Reading and Math Growth and Achievement Gaps in Grades 1–5: A Cohort-Sequential Longitudinal Approach'. *Journal of Educational Psychology*, 112: 4, 718. <https://doi.org/10.1037/edu0000400>
- Soe, K., Koki, S. and Chang, J.M. (2000) 'Effect of Computer-Assisted Instruction (CAI) on Reading Achievement: A Meta-Analysis'. Honolulu, HI: Pacific Resources Education and Learning.
- Sparks, R.L., Patton, P. and Murdoch, A. (2014) 'Early Reading Success and Its Relationship to Reading Achievement and Reading Volume: Replication of '10 Years Later''. *Reading and Writing*, 27: 1, 189–211. <https://doi.org/10.1007/s11145-013-9439-2>
- Speciani, E.R., Merewood, J., Oades, F. and Subosa, M. (2025) '*Independent Evaluation of the Lexia® Core5® Reading Programme: A Cluster Randomised Controlled Trial: Statistical Analysis Plan*'. London: Education Endowment Foundation. Available at: <https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/lexia-core5-reading-2022-2023-trial> (accessed 20 10 2025).
- Sterne, J.A.C., White, I.R., Carlin, J.B., Spratt, M., Royston, P., Kenward, M.G., Wood, A.M. and Carpenter, J.R. (2009) 'Multiple Imputation for Missing Data in Epidemiological and Clinical Research: Potentials and Pitfalls'. *British Medical Journal*, 338. <https://doi.org/10.1136/bmj.b2393>

- Tohara, A.J.T., Shuhidan, M.S., Bahry, F.D.S. and bin Nordin, M.N. (2021) 'Exploring Digital Literacy Strategies for Students With Special Educational Needs in the Digital Age'. Turkish *Journal of Computer and Mathematics Education*, 12: 9, 3345–3358. <https://doi.org/10.17762/turcomat.v12i9.5741>
- Tracey, L., Elliott, L., Fairhurst, C., Mandefield, L., Fountain, I. and Ellison, S. (2022) '*Lexia Reading Core5®: Evaluation Report*'. London: Education Endowment Foundation. Available at: <https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/lexia> (accessed 20 10 2025).
- Van Der Stuyf, R.R. (2002) '*Scaffolding as a Teaching Strategy*'. Adolescent Learning and Development. Available at: https://ateachingpath1.weebly.com/uploads/1/7/8/9/17892507/stuyf_2002.pdf (accessed 13 09 2026).
- Vygotsky, L. (1978) '*Mind in Society*'. Cambridge, MA: Harvard University Press.
- Wanzek, J., Stevens, E.A., Williams, K.J., Scammacca, N., Vaughn, S. and Sargent, K. (2018) 'Current Evidence on the Effects of Intensive Early Reading Interventions'. *Journal of Learning Disabilities*, 51: 6, 612–624. <https://doi.org/10.1177/0022219418775110>
- Watson, S. (2024) '*Package 'crctStepdown': Univariate Analysis of Cluster Trials with Multiple Outcomes*'. Available at: <https://cran.r-project.org/web/packages/crctStepdown/crctStepdown.pdf> (accessed 23 01 2025).
- Woodcock, R.W. (2011) '*Woodcock Reading Mastery Tests, Third Edition WRMT-III*'. London: Pearson. Available at: www.pearsonclinical.co.uk/en-gb/Store/Professional-Assessments/Academic-Learning/Woodcock-Reading-Mastery-Tests%2C-Third-Edition/p/P100009264 (accessed 20 10 2025).

Appendix A: The EEF cost rating

Appendix A Table 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

OUTCOME: Reading proficiency (measure – Woodcock Reading Mastery Test, Third Edition [WRMT-III]; Woodcock, 2011)

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		Risk rating	Comments				
Threat 1: Confounding		Low	Randomisation was conducted in Stata by a member of the evaluation team who was blind to treatment. Treatment and control were well balanced on FSM, EAL, SEND, geographic characteristics, and prior attainment.				

Threat 2: Concurrent interventions	Moderate	Schools in both treatment and control groups were delivering their usual literacy provision and targeted reading support alongside the trial. Participating schools used other reading interventions and technology-supported instruction as part of business as usual. No indication that concurrent interventions were systematically imbalanced across groups, but this was not controlled for analytically.
Threat 3: Experimental effects	Low	The evaluator engaged with the possibility of Hawthorne effects noting that there was limited evidence with treatment schools delivering reading sessions more formally and consistently than control schools. There is little evidence of John Henry effects.
Threat 4: Implementation fidelity	Low	The intervention was delivered with fidelity and compliance was high.
Threat 5: Missing data	Low	Missing data was explored in line with the EEF statistical analysis guidance (EEF, 2022b) using a logistic model and multiple imputation. Missingness was not differential by treatment group, and the multiple imputation analysis produced a similar estimate (0.138) to the complete case analysis (0.145).
Threat 6: Measurement outcomes	Low	The outcome measure is the WRMT-III, a standardised reading assessment with strong psychometric properties. Oral Reading Fluency was removed from the primary composite due to concerns around it being distressing for pupils.
Threat 7: Selective reporting	Low	The study is registered, and the protocol and SAP were published ahead of this report and followed (Merewood <i>et al.</i> , 2025; Speciani <i>et al.</i> , 2025). Deviations from the original plans have been clearly presented (e.g. the use of three subtests rather than four subtests).

- **Initial padlock score:** 4 padlocks – An intact RCT, sufficiently powered to 0.116 but there was attrition at the pupil level between randomisation and analysis (12.93%) which loses the trial 1 padlock.
- **Reason for adjustment for threats to validity:** [0] – No adjustments as there is no evidence of serious threats to the internal validity of the study.
- **Final padlock score (initial score adjusted for threats to validity):** 4 padlocks.

Appendix C: Changes since the previous evaluation

Appendix C Table 1: Changes since the previous evaluation

	Feature	Efficacy to effectiveness stage
Intervention	Intervention content	Not applicable
	Delivery model	Not applicable
	Intervention duration	Not applicable
Evaluation	Eligibility criteria	Efficacy: School has approximately 50 pupils per year group Effectiveness: School has between 8 and 16 Year 2 pupils who would be eligible for Lexia Efficacy: School is not currently involved in any other trial by the EEF focusing on Key Stage 1 literacy or aiming to achieve change at a whole-school level Effectiveness: School is not currently involved in any other trial by the EEF focusing on Year 2
	Level of randomisation	Efficacy: Pupil-level randomisation Effectiveness: School-level randomisation
	Outcomes and baseline	Efficacy: Primary outcome was WRMT-III composite of the scores of four subtests: Word Identification; Word Attack; Passage Comprehension; and Oral Reading Fluency Effectiveness: Primary outcome was WRMT-III composite of the scores of three subtests: Word Identification; Word Attack; and Passage Comprehension
	Control condition	Efficacy: Year 2 pupils not randomised to receive the intervention Effectiveness: Schools not randomised to receive the intervention

Appendix D: Effect size estimation and other tables

Appendix D Table 1: Primary model regression output

	Coefficient (SD)
Intercept	25.981*** (5.306)
Treatment	5.346*** (1.251)
Baseline score	0.855*** (0.015)
EIA	0.573 (1.256)
Cohort	0.553 (1.254)

*** P -value < 0.01

Estimation of effect sizes

Appendix D Table 2: Effect size estimation primary model

Outcome	Unadjusted differences in means	Adjusted differences in means	Intervention group		Control group		Pooled variance
			N (missing)	Variance of outcome	N (missing)	Variance of outcome	
WRMT-III Composite	4.158	5.242	1,069 (18)	1,431.350	1,128 (16)	1,205.436	1,315.357

Appendix D Table 3: Effect size estimation secondary model

Outcome	Unadjusted differences in means	Adjusted differences in means	Intervention group		Control group		Pooled variance
			N (missing)	Variance of outcome	N (missing)	Variance of outcome	
Word Identification	0.971	1.210	1,092 (1)	255.760	1,146 (2)	215.628	235.210
Word Attack	1.709	2.304	1,081 (11)	194.480	1,145 (6)	154.992	174.168
Passage Comprehension	1.470	1.620	1,079 (14)	149.843	1,138 (11)	147.830	148.810
Oral Reading Fluency	0.081	0.359	972 (14)	212.502	1,024 (13)	187.922	199.892

Appendix D Table 4: Effect size estimation compliance models

Outcome	Unadjusted differences in means	Adjusted differences in means	Intervention group		Control group		Pooled variance
			N (missing)	Variance of outcome	N (missing)	Variance of outcome	
WRMT-III Composite (compliance analysis)	4.221	5.425	1,059 (18)	1,434.382	1,128 (16)	1,205.436	1,316.294
WRMT-III Composite (dosage analysis)	4.221	0.058	1,059 (18)	1,434.382	1,128 (16)	1,205.436	1,316.294
WRMT-III Composite (2x Lexias)	4.221	5.735	1,059 (18)	1,434.382	1,128 (16)	1,205.436	1,316.294
WRMT-III Composite (3x Lexias)	4.221	7.317	1,059 (18)	1,434.382	1,128 (16)	1,205.436	1,316.294
WRMT-III Composite (4x Lexias)	4.221	13.062	1,059 (18)	1,434.382	1,128 (16)	1,205.436	1,316.294
WRMT-III Composite (5x Lexias)	4.221	27.574	1,059 (18)	1,434.382	1,128 (16)	1,205.436	1,316.294

Appendix D Table 5: Effect size estimation missing data analysis

Outcome	Unadjusted differences in means	Adjusted differences in means	Intervention group		Control group		Pooled variance
			N (missing)	Variance of outcome	N (missing)	Variance of outcome	
WRMT-III Composite	10.5339	5.5642	1,087	1,719.681	1,142	1,553.678	1,634.6293

Appendix D Table 6: Effect size estimation FSM subsample analysis

Outcome	Unadjusted differences in means	Adjusted differences in means	Intervention group		Control group		Pooled variance
			N (missing)	Variance of outcome	N (missing)	Variance of outcome	
WRMT-III Composite	4.7002	6.3296	364 (5)	1,551.434	360 (7)	1,195.417	1,374.4117

Appendix D Table 7: Effect size estimation FSM interaction term analysis

Outcome	Unadjusted differences in means	Treatment coefficient	Interaction coefficient	Intervention group		Control group		Pooled variance
				N (missing)	Variance of outcome	N (missing)	Variance of outcome	
WRMT-III Composite	4.2422	4.8269	1.4152	1,069 (18)	1,431.35	1,126 (16)	1,203.199	1,314.3097

Appendix D Table 8: Effect size estimation EAL subsample analysis

Outcome	Unadjusted differences in means	Adjusted differences in means	Intervention group		Control group		Pooled variance
			N (missing)	Variance of outcome	N (missing)	Variance of outcome	
WRMT-III Composite	10.6421	7.774552	230 (4)	1,333.086	254 (1)	1,092.981	1,207.0559

Appendix D Table 9: Effect size estimation EAL interaction term analysis

Outcome	Unadjusted differences in means	Treatment coefficient	Interaction coefficient	Intervention group		Control group		Pooled variance
				N (missing)	Variance of outcome	N (missing)	Variance of outcome	
WRMT-III Composite	4.0179	4.8741	2.0686	1,068 (19)	1,430.697	1,125 (17)	1,203.864	1,314.3297

Appendix D Table 10: Effect size estimation SEND subsample analysis

Outcome	Unadjusted differences in means	Adjusted differences in means	Intervention group		Control group		Pooled variance
			N (missing)	Variance of outcome	N (missing)	Variance of outcome	
WRMT-III Composite	0.316	1.605841	291 (5)	1,710.507	318 (5)	1,352.243	1,523.4074

Appendix D Table 11: Effect size estimation SEND interaction term analysis

Outcome	Unadjusted differences in means	Treatment coefficient	Interaction coefficient	Intervention group		Control group		Pooled variance
				N (missing)	Variance of outcome	N (missing)	Variance of outcome	
WRMT-III Composite	4.2422	6.6711	-5.2542	1,069 (18)	1,431.35	1,126 (16)	1,203.199	1,314.3097

Appendix D Table 12: Effect size estimation baseline reading score subgroup models

Outcome	Unadjusted differences in means	Adjusted differences in means	Intervention group		Control group		Pooled variance
			N (missing)	Variance of outcome	N (missing)	Variance of outcome	
WRMT-III Composite (Bottom 20% subgroup)	0.797	0.283	44 (0)	919.608	27 (0)	895.251	910.430
WRMT-III Composite (Bottom 30% subgroup)	4.801	6.381	127 (0)	1,027.910	104 (0)	1,025.173	1,026.679
WRMT-III Composite (Bottom 40% subgroup)	0.467	3.837	277 (0)	1,118.676	284 (0)	933.086	1,024.719
WRMT-III Composite (Bottom 50% subgroup)	1.553	3.727	558 (0)	1,049.020	583 (0)	934.234	990.367
WRMT-III Composite (Top 50% subgroup)	7.456	7.017	545 (0)	642.063	511 (0)	576.157	608.047
WRMT-III Composite (Top 40% subgroup)	7.461	7.474	259 (0)	481.980	274 (0)	421.061	450.660
WRMT-III Composite (Top 30% subgroup)	5.777	7.836	85 (0)	474.919	93 (0)	452.449	463.173

Appendix D Table 13: Effect size estimation cohort models

Outcome	Unadjusted differences in means	Adjusted differences in means	Intervention group		Control group		Pooled variance
			N (missing)	Variance of outcome	N (missing)	Variance of outcome	
WRMT-III Composite (Cohort 1)	1.914	4.295	550 (15)	1,522.798	592 (11)	1,157.129	1,333.227
WRMT-III Composite (Cohort 2)	6.528	6.211	519 (3)	1,324.176	536 (5)	1,260.963	1,292.06

Appendix D Table 14: Effect size estimation Oral Reading Fluency model

Outcome	Unadjusted differences in means	Adjusted differences in means	Intervention group		Control group		Pooled variance
			N (missing)	Variance of outcome	N (missing)	Variance of outcome	
WRMT-III Composite plus Oral Reading Fluency	4.861	5.864	965 (14)	2,096.098	1,017 (13)	1,766.729	1,927.089

Logit model of missingness

Appendix D Table 15: Logit model results

	Effect on outcome missing	Standard error	P-value	95% CI
Treatment	0.0169	0.0140	0.228	(-0.0106, 0.0444)
Gender	0.0073	0.0129	0.569	(-0.0179, 0.0326)
FSM	0.0330	0.0146	0.024	(0.0043, 0.0616)
EAL	0.0266	0.0161	0.098	(-0.0049, 0.0581)
SEND	0.0144	0.0149	0.336	(-0.0150, 0.0436)
Baseline score	-0.0009	0.0002	0.000	(-0.0013, -0.0005)
Urban	-0.0083	0.0199	0.675	(-0.0474, 0.0307)
EIA	-0.0064	0.0143	0.657	(-0.0345, 0.0217)
'Inadequate' Ofsted rating	0.1796	0.0451	0.000	(0.0913, 0.2679)
'Outstanding' Ofsted rating	-0.0129	0.0256	0.615	(-0.0630, 0.0373)
'Requires improvement' Ofsted rating	0.0368	0.0289	0.203	(-0.0198, 0.0934)
FSM proportion	0.0265	0.0504	0.599	(-0.0723, 0.1254)

Results from first stage of CACE analysis

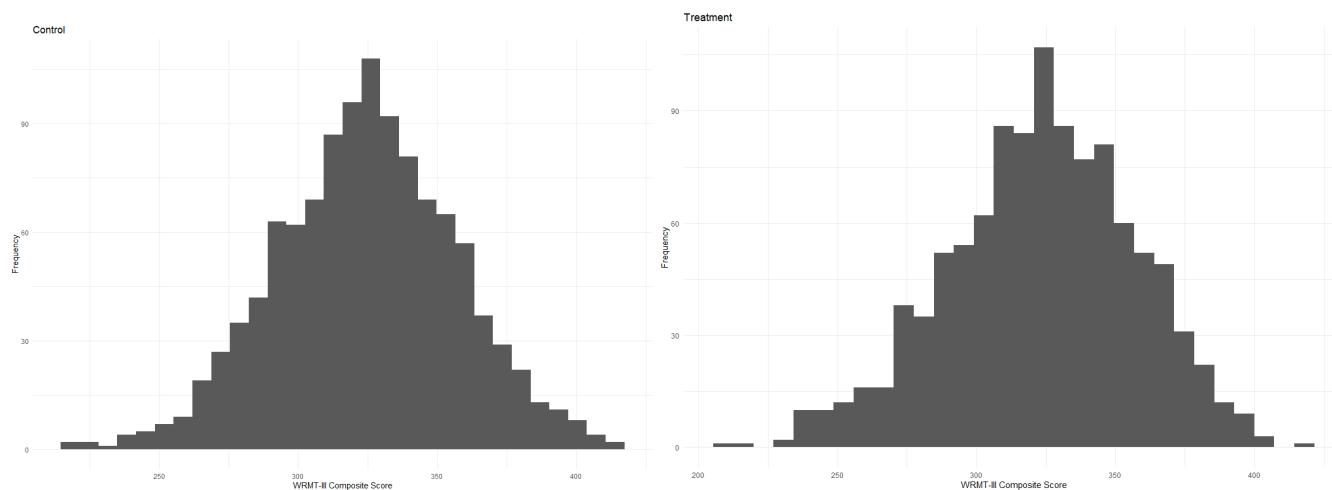
Appendix D Table 16: Results of first stage regression in CACE analysis, by compliance measure

	Coefficient (SD)					
	Compliance model	Dosage model	2x Lexias	3x Lexias	4x Lexias	5x Lexias
Intercept	-0.047 (0.027)	-7.324 (5.750)	-0.083 (0.047)	-0.126 (0.077)	-0.006 (0.084)	-0.024 (0.067)
Treatment	0.974*** (0.008)	91.196*** (2.773)	0.921*** (0.017)	0.722*** (0.033)	0.404*** (0.040)	0.192*** (0.033)
Baseline score	0.000 (0.000)	0.005 (0.010)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
EIA	-0.003 (0.008)	1.315 (2.780)	-0.008 (0.017)	0.006 (0.033)	0.010 (0.040)	0.019 (0.033)
Cohort	0.020* (0.008)	3.274 (2.778)	0.034* (0.017)	0.060 (0.033)	-0.004 (0.040)	0.011 (0.033)

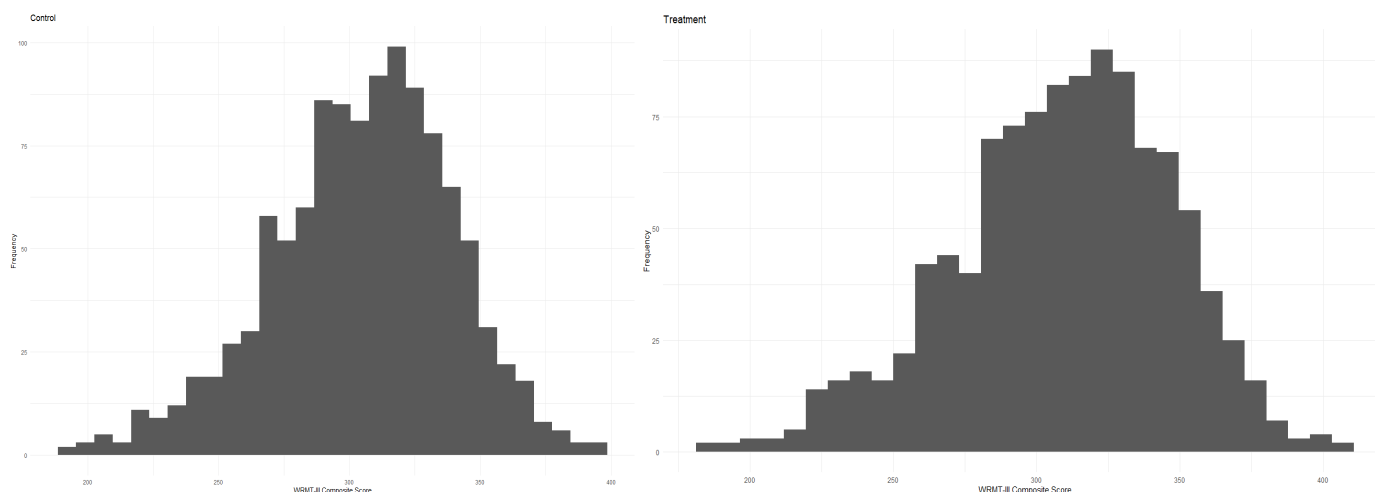
*** p -value < 0.01

Appendix E: Outcome distributions

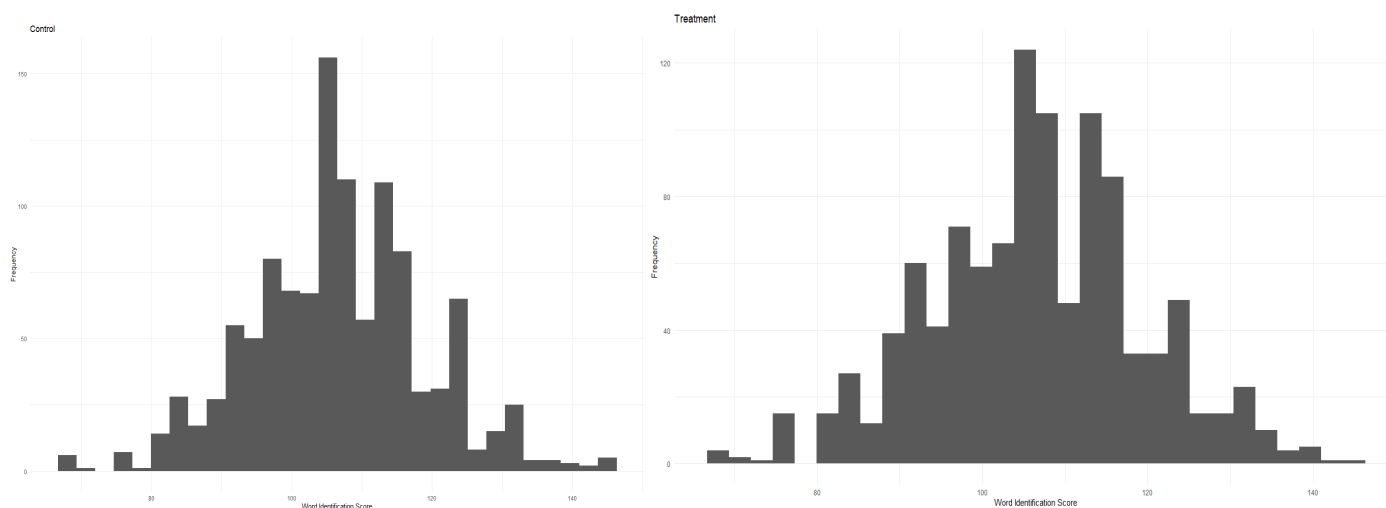
Appendix E Figure 1: Baseline distribution of WRMT-III composite outcome in analytical sample, by treatment arm



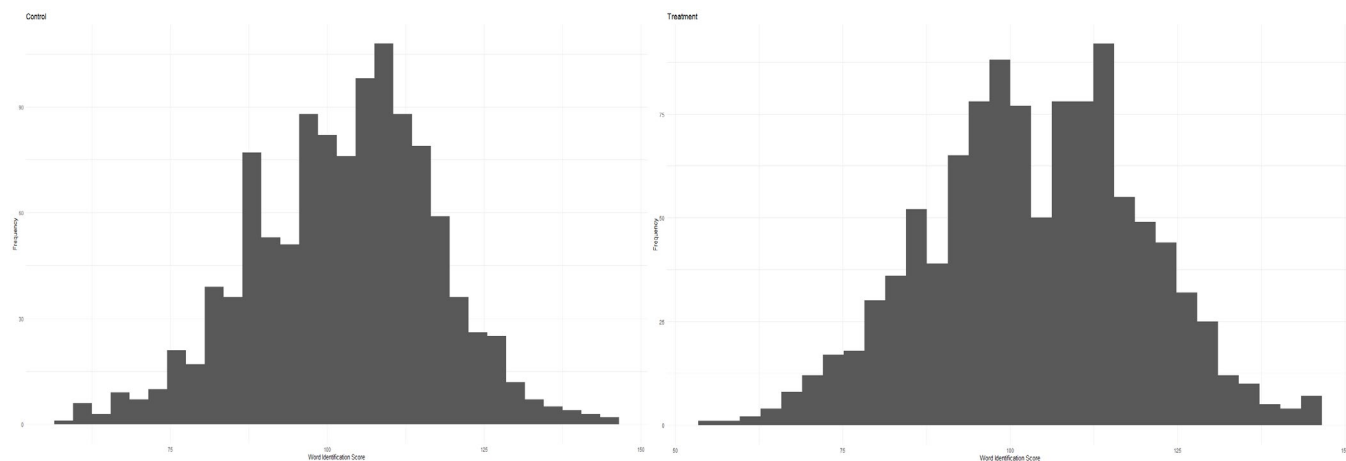
Appendix E Figure 2: Endline distribution of WRMT-III composite outcome in analytical sample, by treatment arm



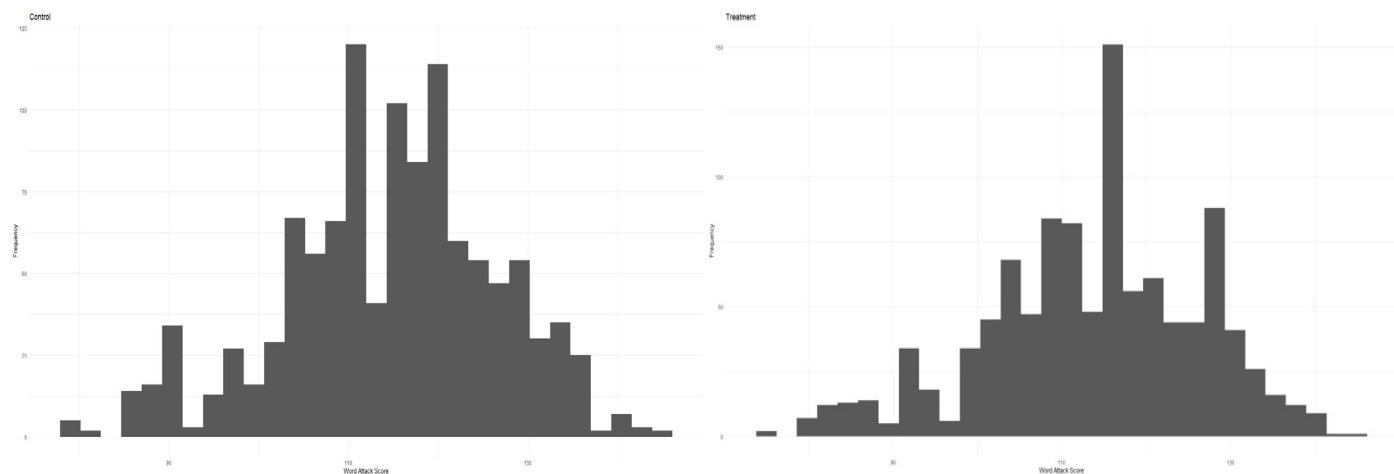
Appendix E Figure 3: Baseline distribution of Word Identification outcome in analytical sample, by treatment arm



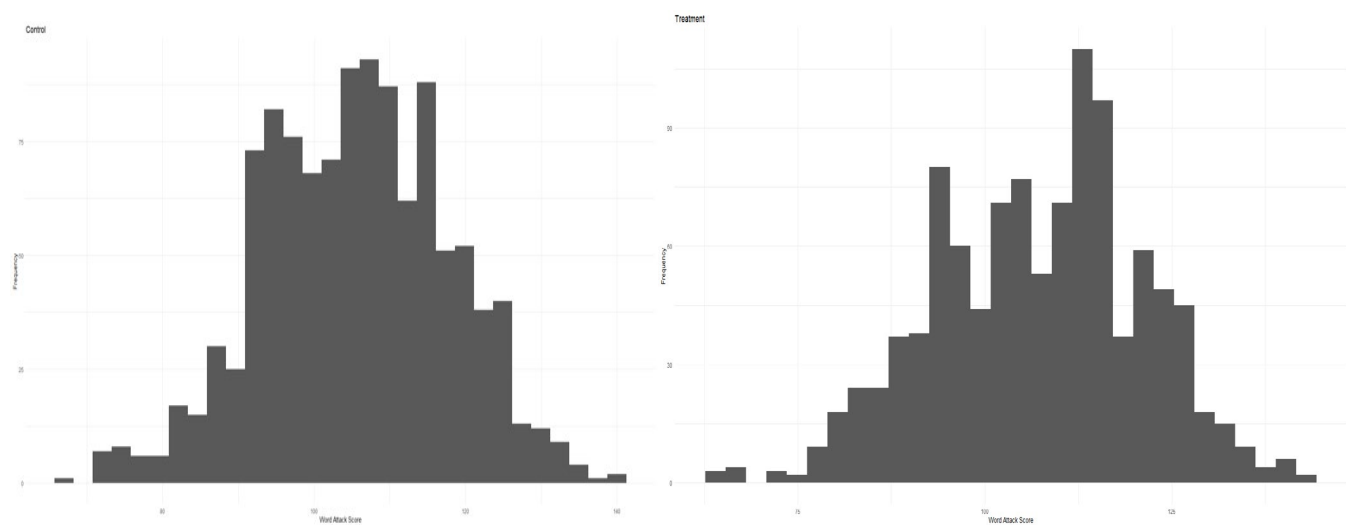
Appendix E Figure 4: Endline distribution of Word Identification outcome in analytical sample, by treatment arm



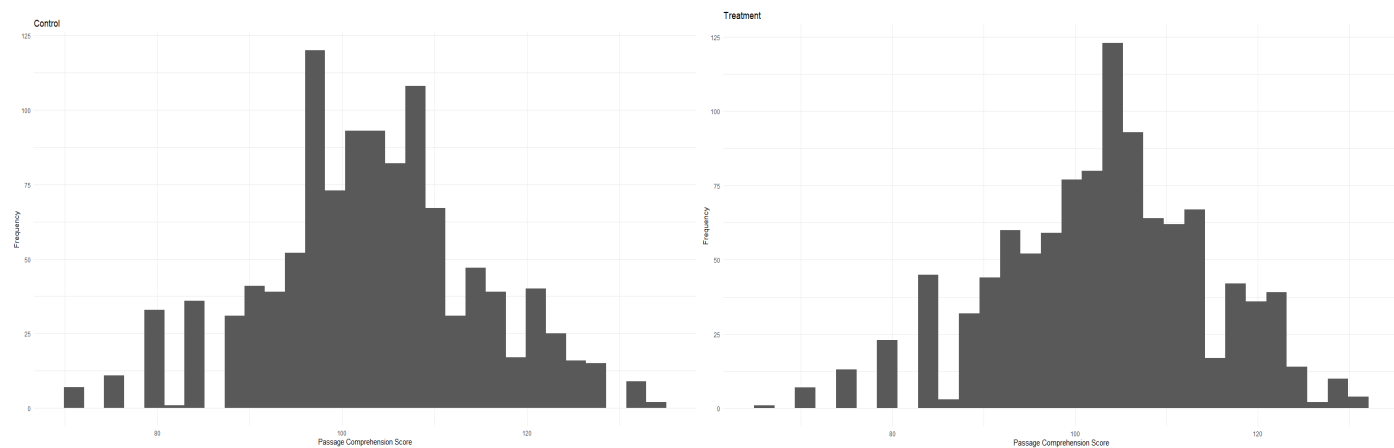
Appendix E Figure 5: Baseline distribution of Word Attack outcome in analytical sample, by treatment arm



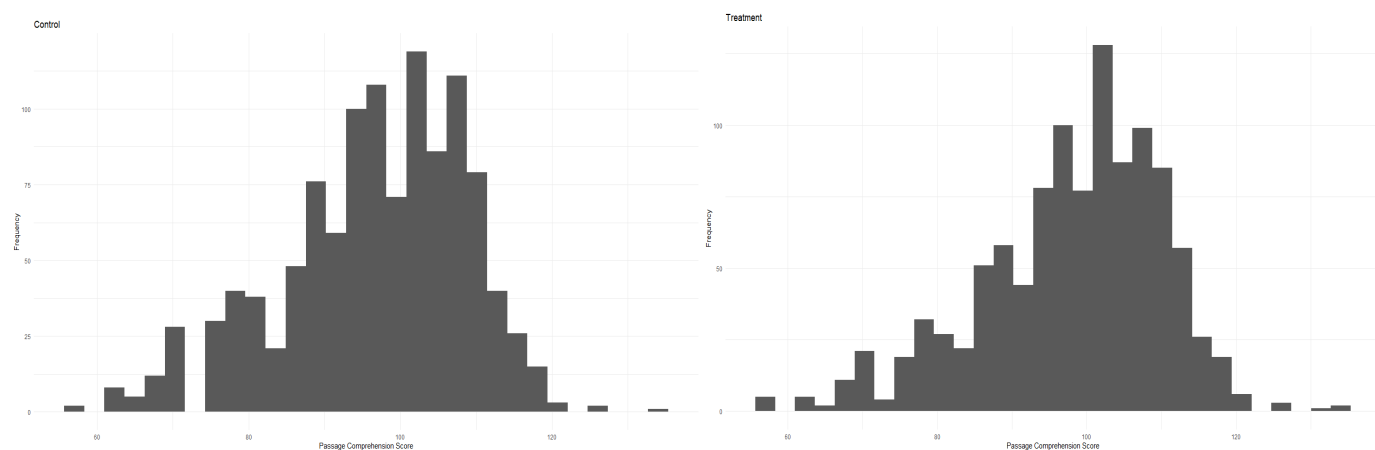
Appendix E Figure 6: Endline distribution of Word Attack outcome in analytical sample, by treatment arm



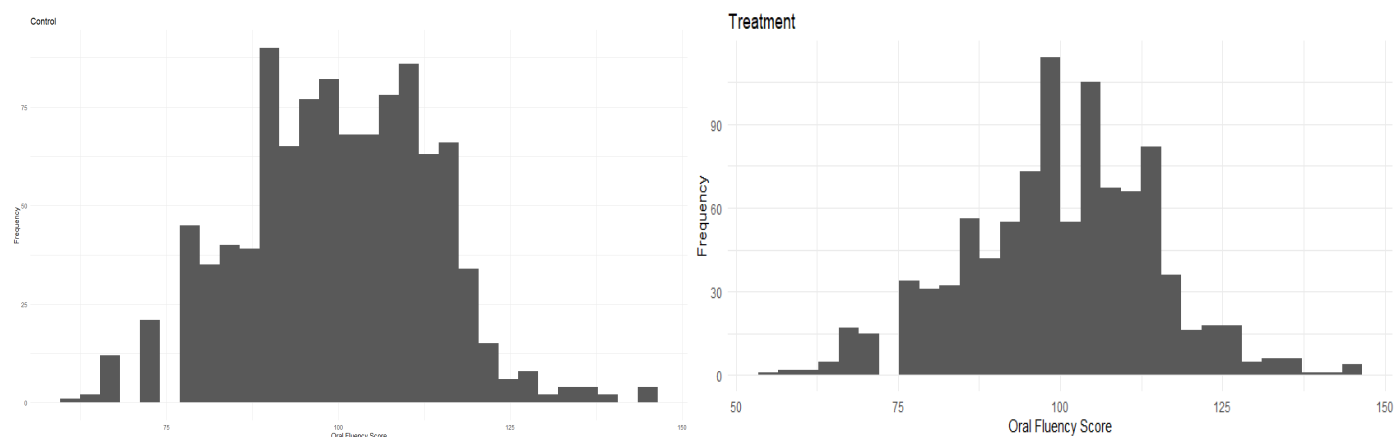
Appendix E Figure 7: Baseline distribution of Passage Comprehension outcome in analytical sample, by treatment arm



Appendix E Figure 8: Endline distribution of Passage Comprehension outcome in analytical sample, by treatment arm

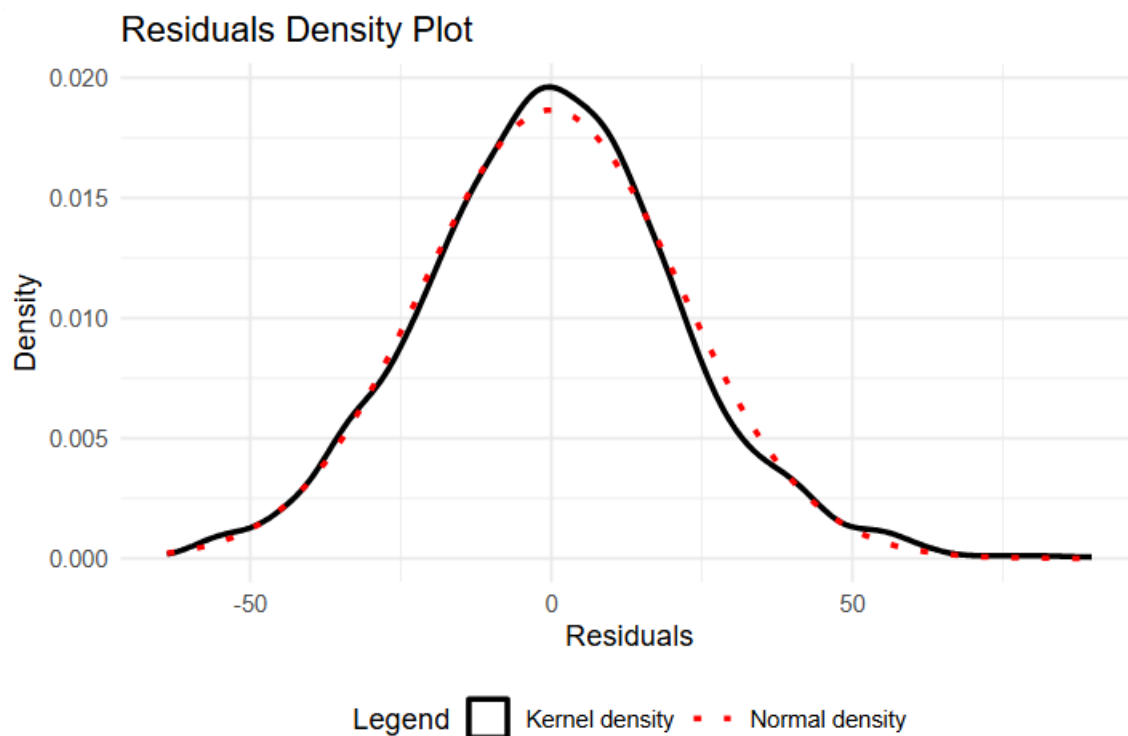


Appendix E Figure 9: Endline distribution of Oral Reading Fluency outcome in analytical sample, by treatment arm

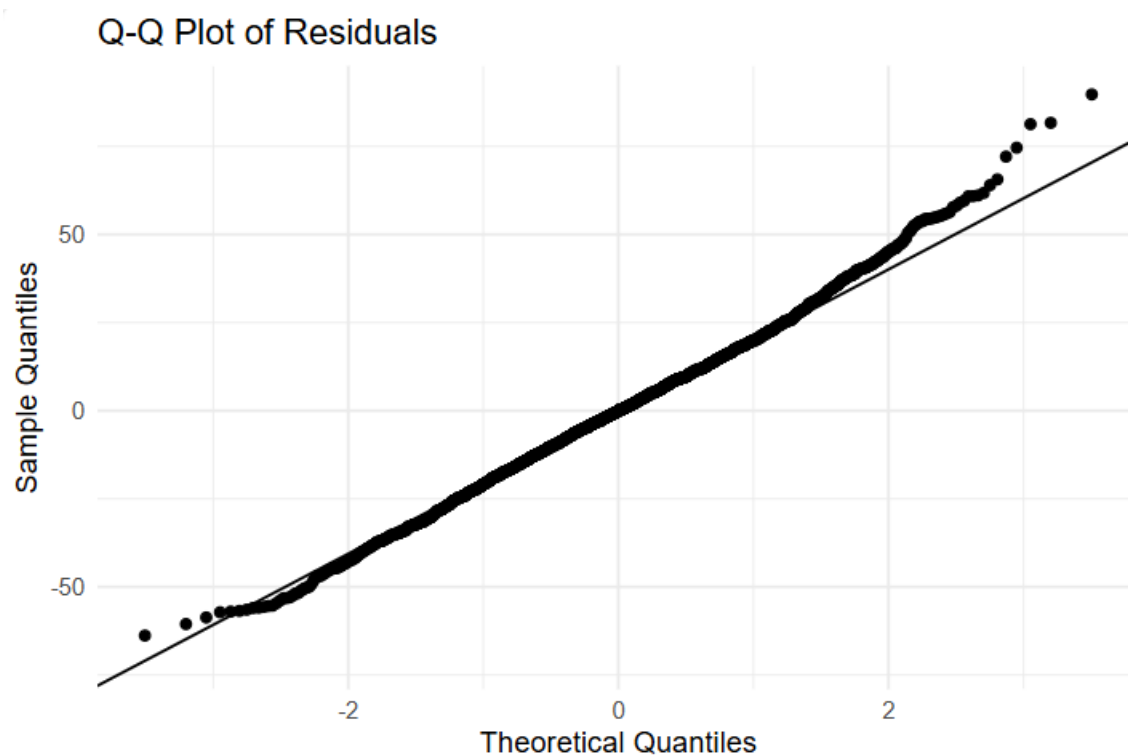


Appendix F: Diagnostic plots

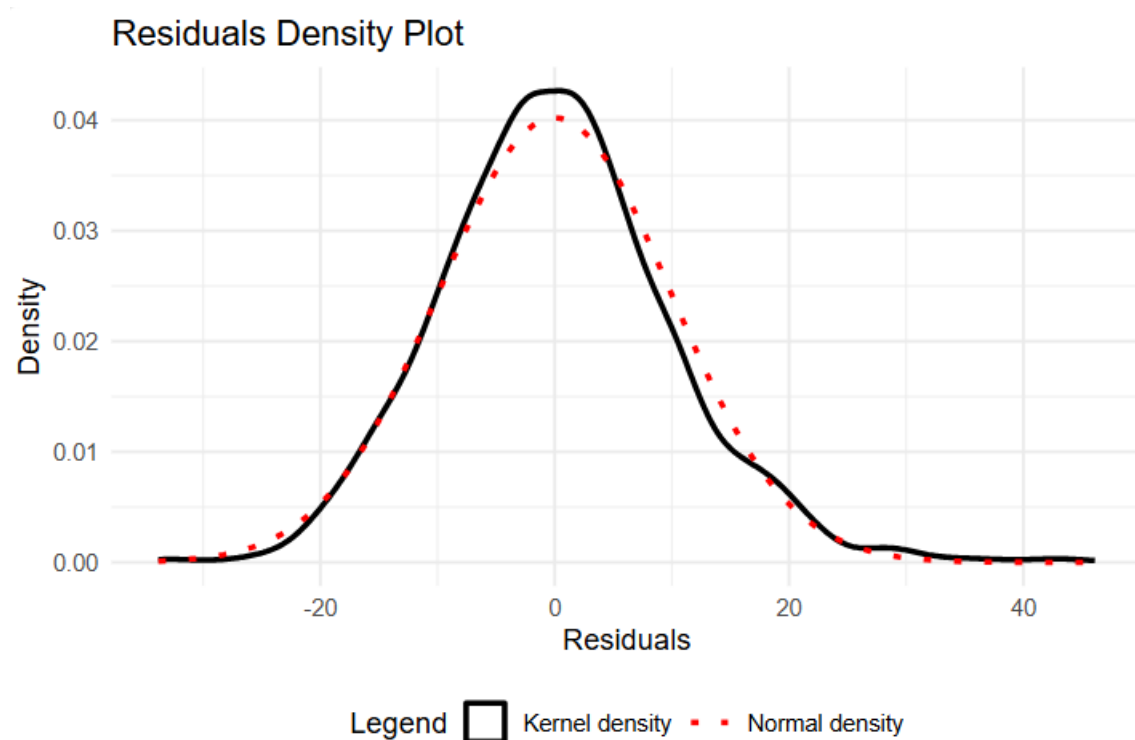
Appendix F Figure 1: Primary analysis residual density plot



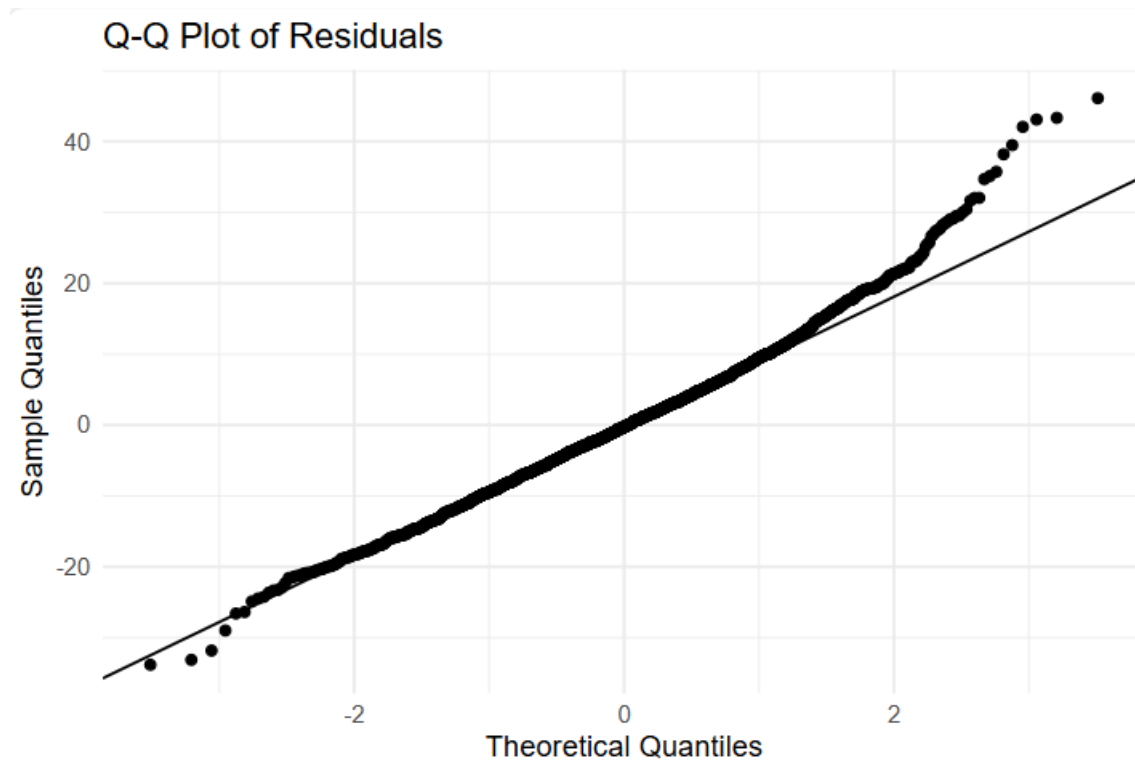
Appendix F Figure 2: Primary analysis residual Q-Q (Quantile-Quantile) plot



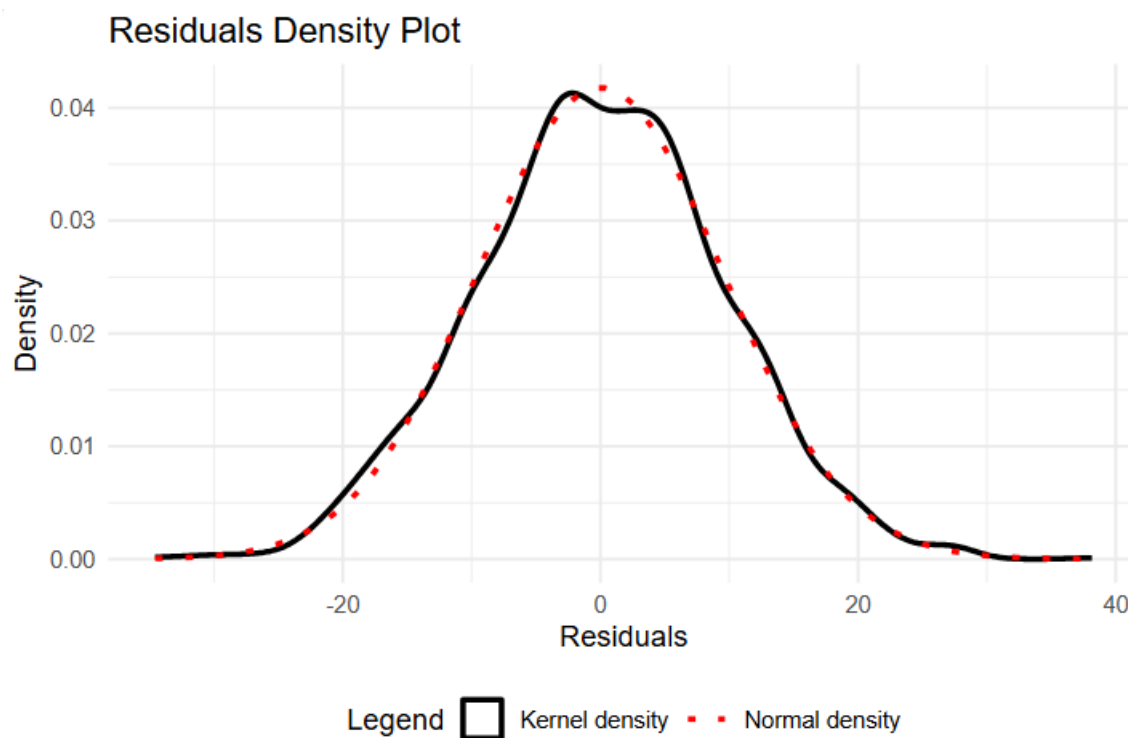
Appendix F Figure 3: Secondary analysis residual density plot (Word Identification)



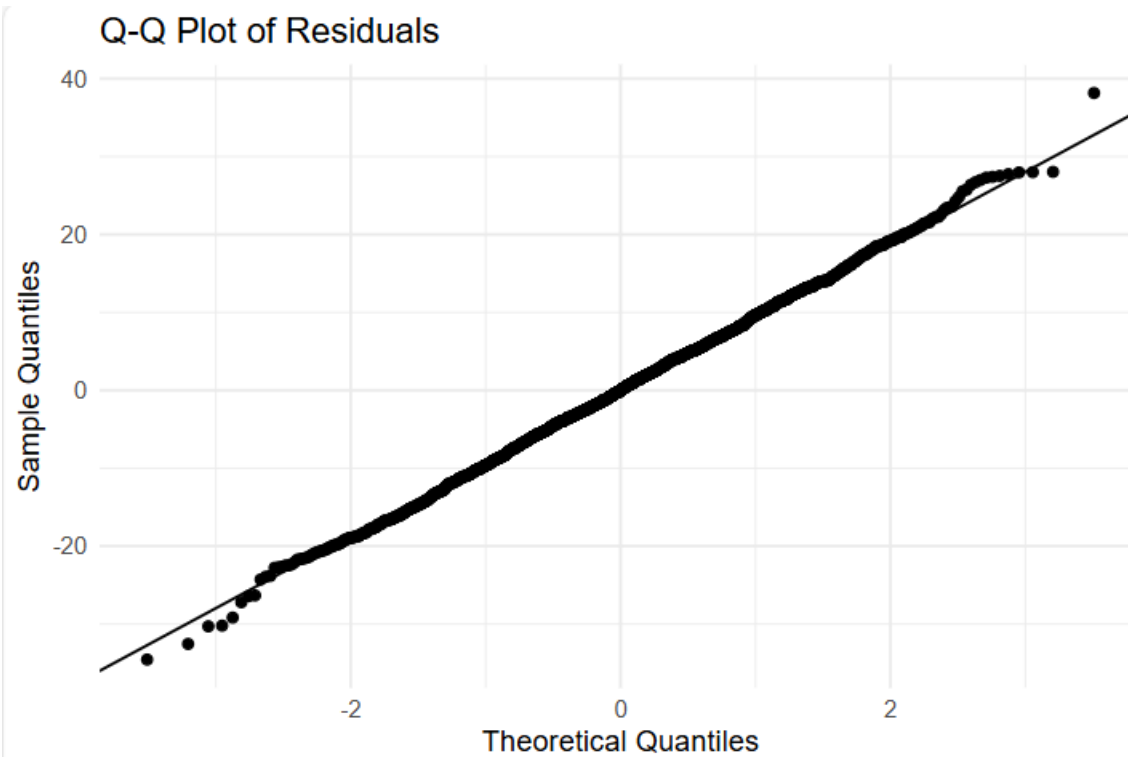
Appendix F Figure 4: Secondary analysis residual Q-Q plot (Word Identification)



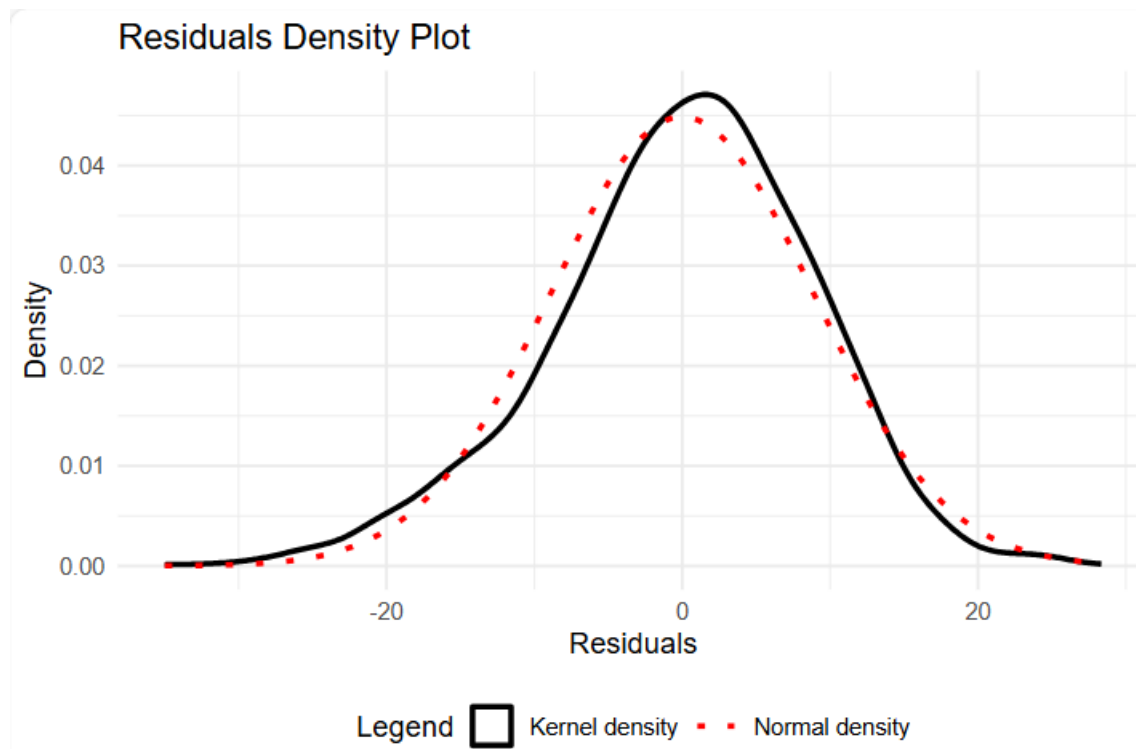
Appendix F Figure 5: Secondary analysis residual density plot (Word Attack)



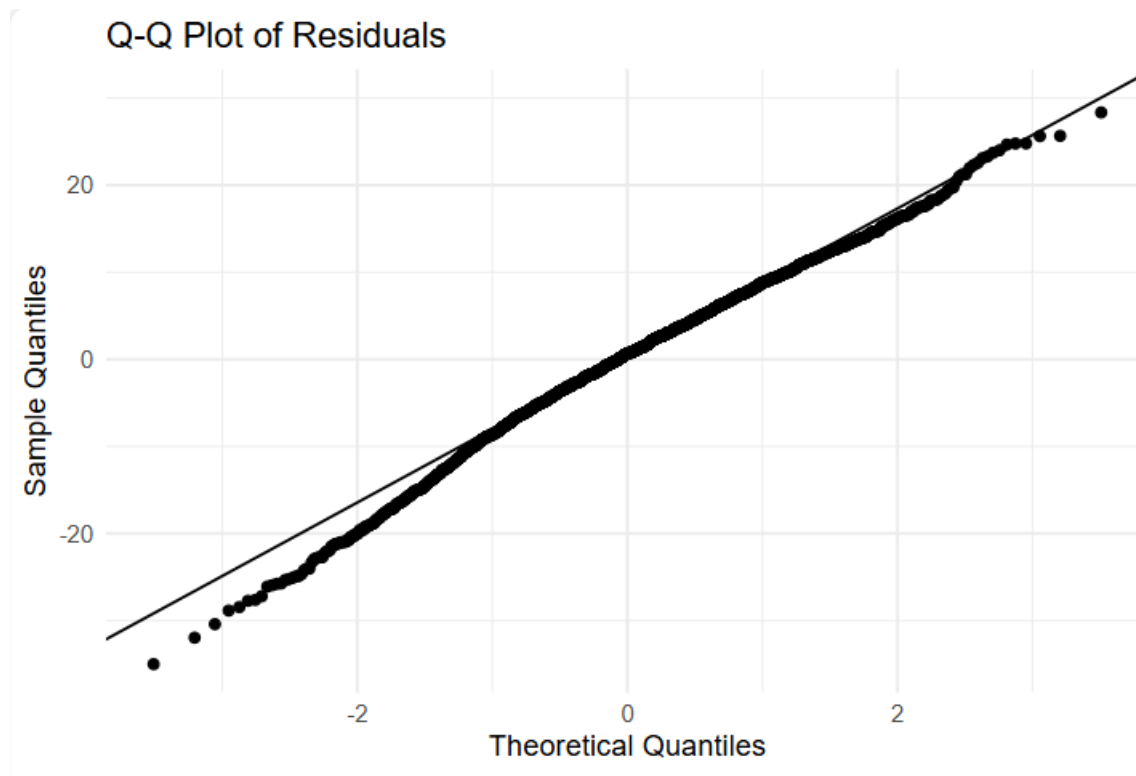
Appendix F Figure 6: Secondary analysis residual Q-Q plot (Word Attack)



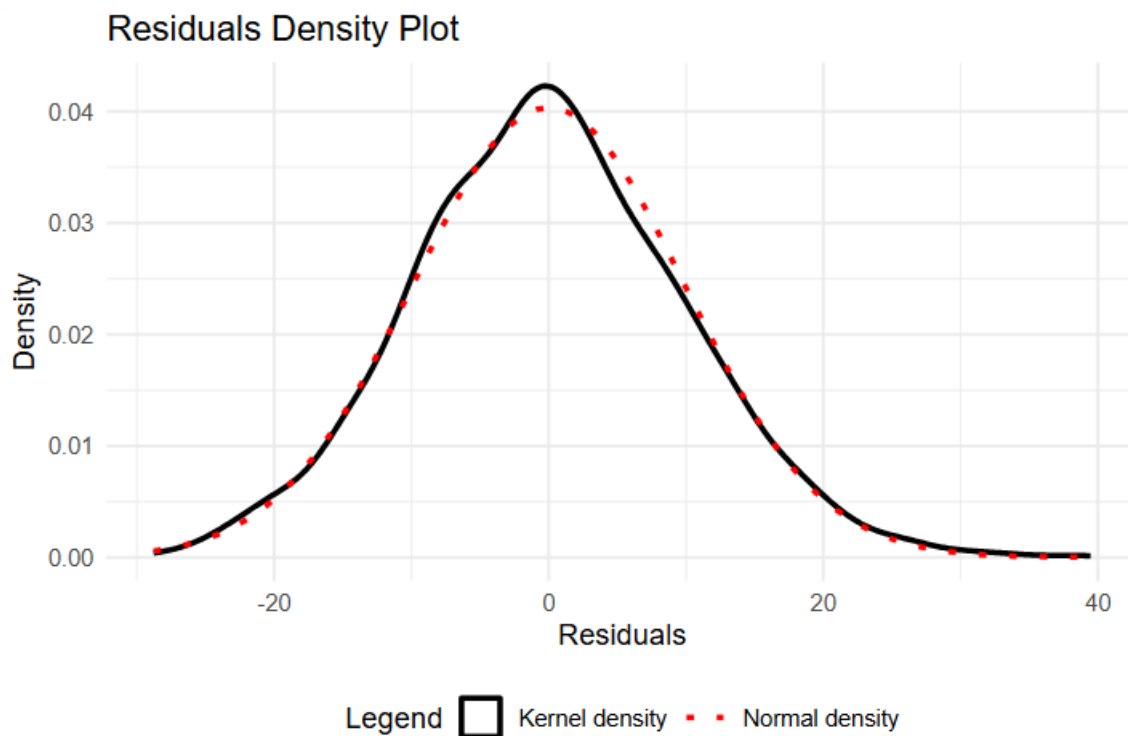
Appendix F Figure 7: Secondary analysis residual density plot (Passage Comprehension)



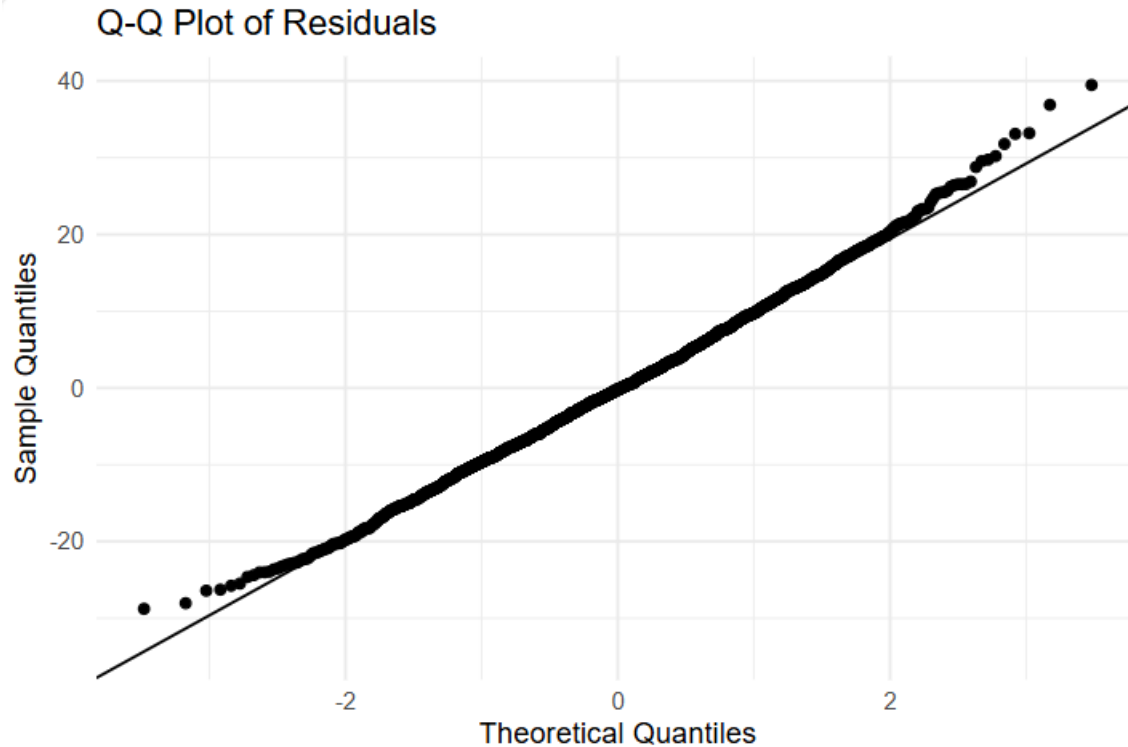
Appendix F Figure 8: Secondary analysis residual Q-Q plot (Passage Comprehension)



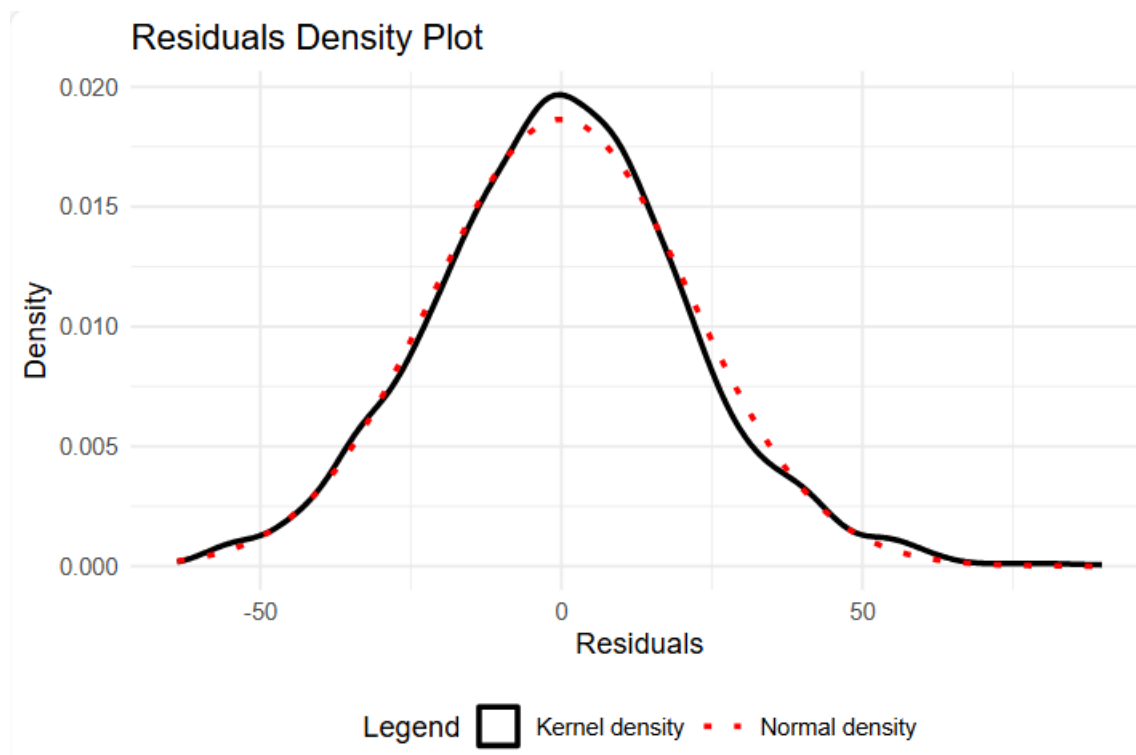
Appendix F Figure 9: Secondary analysis residual density plot (Oral Reading Fluency)



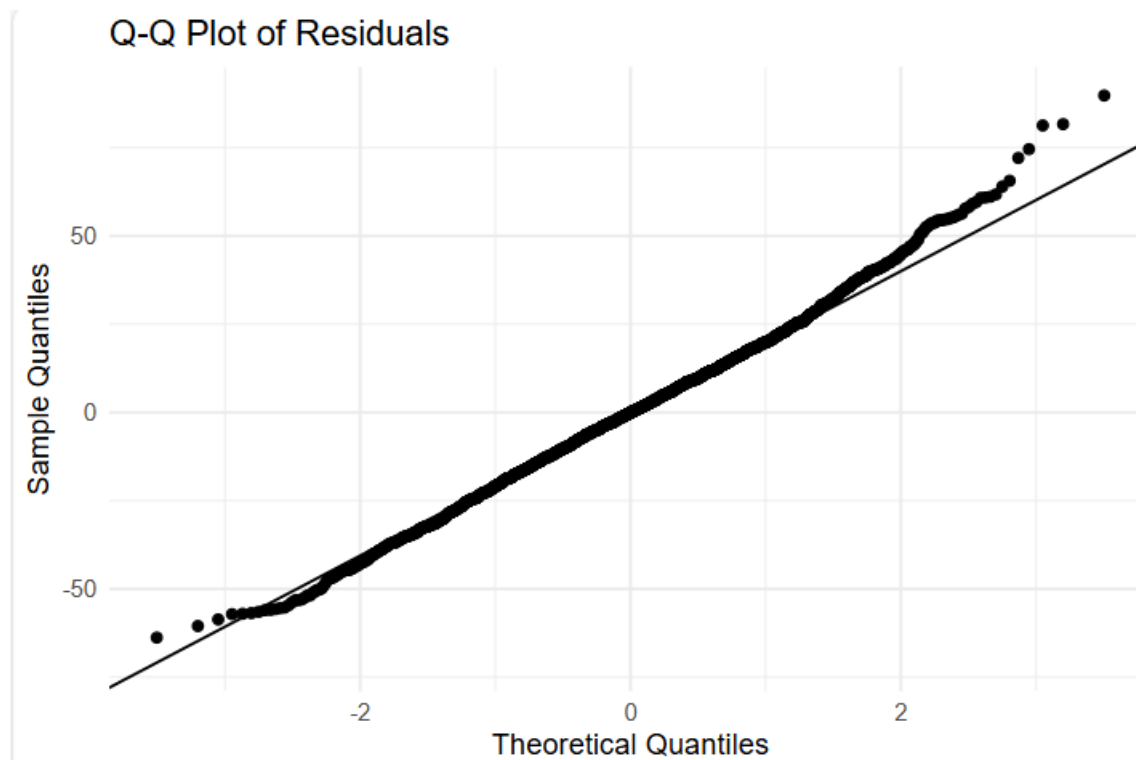
Appendix F Figure 10: Secondary analysis residual Q-Q plot (Oral Reading Fluency)



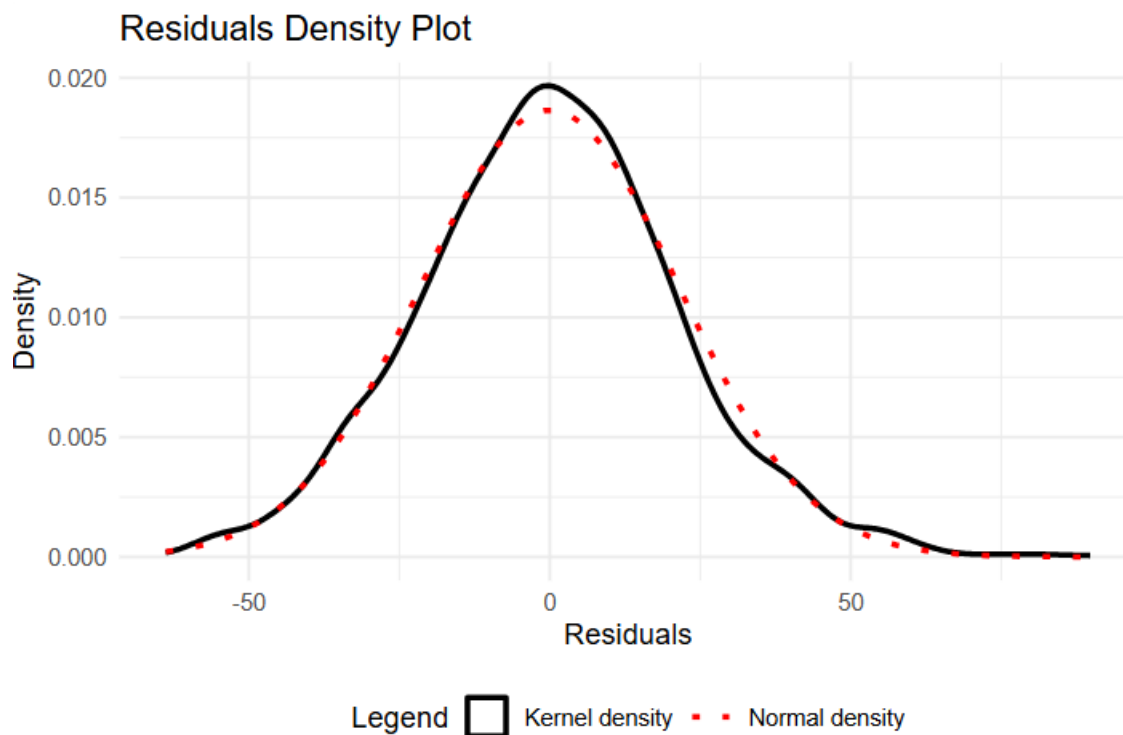
Appendix F Figure 11: Compliance analysis residual density plot



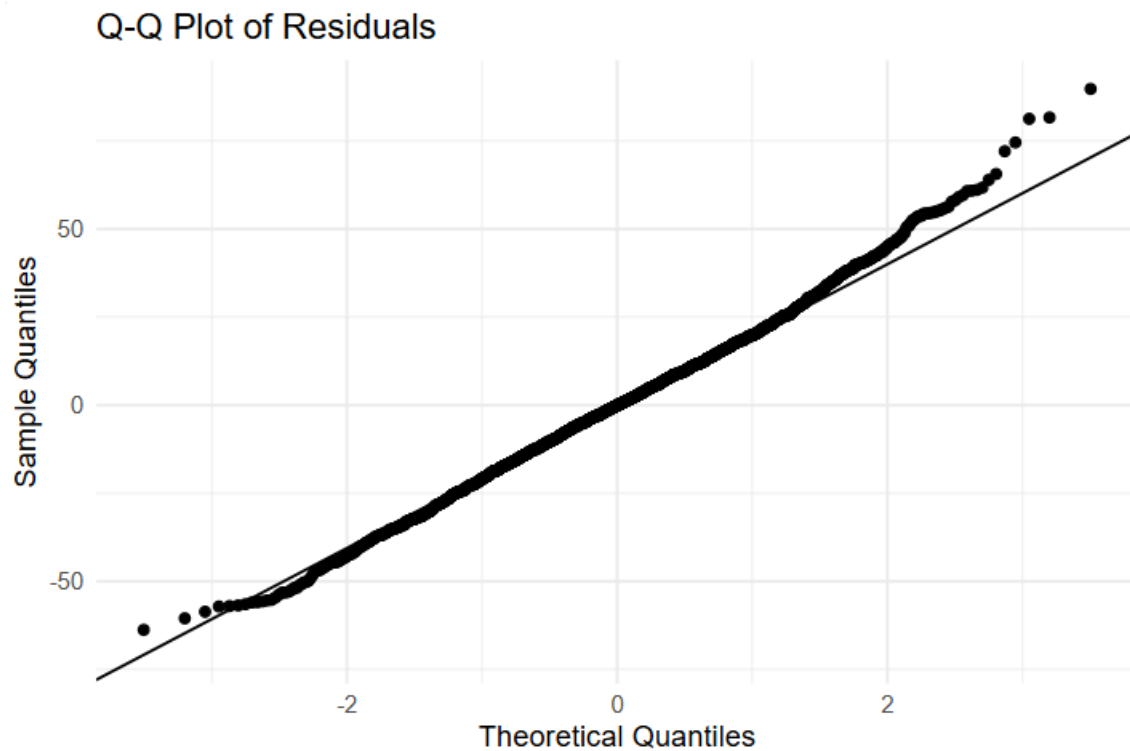
Appendix F Figure 12: Compliance analysis residual Q-Q plot



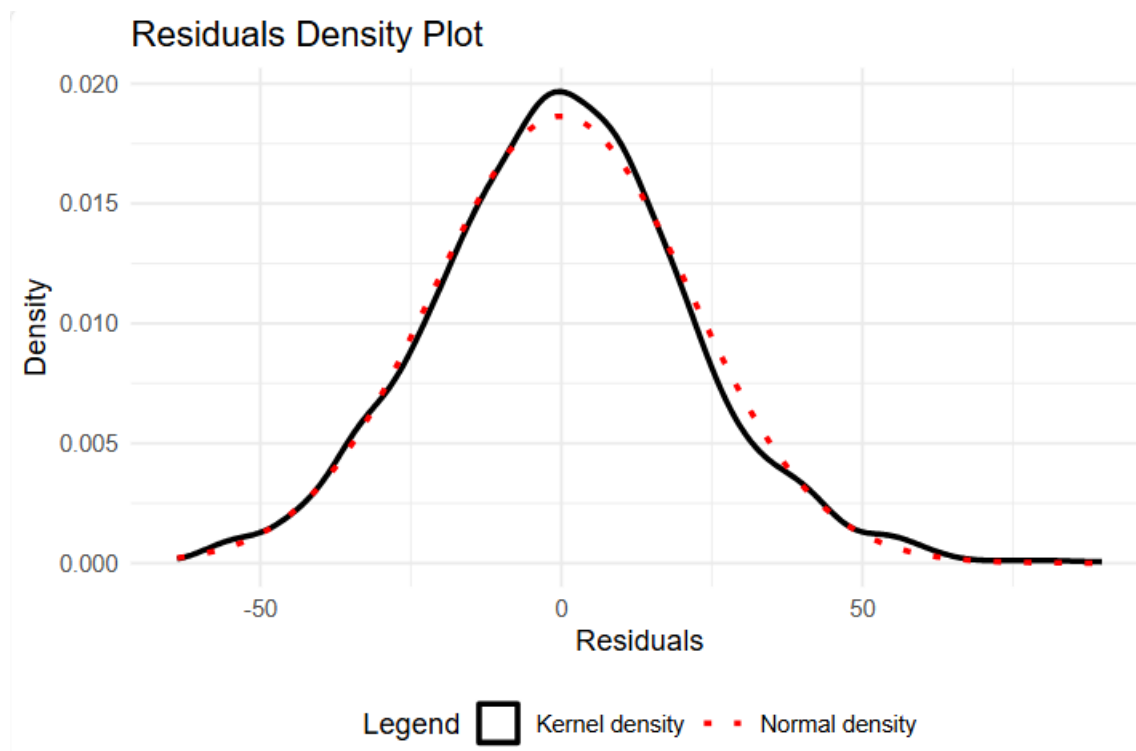
Appendix F Figure 13: Dosage analysis residual density plot



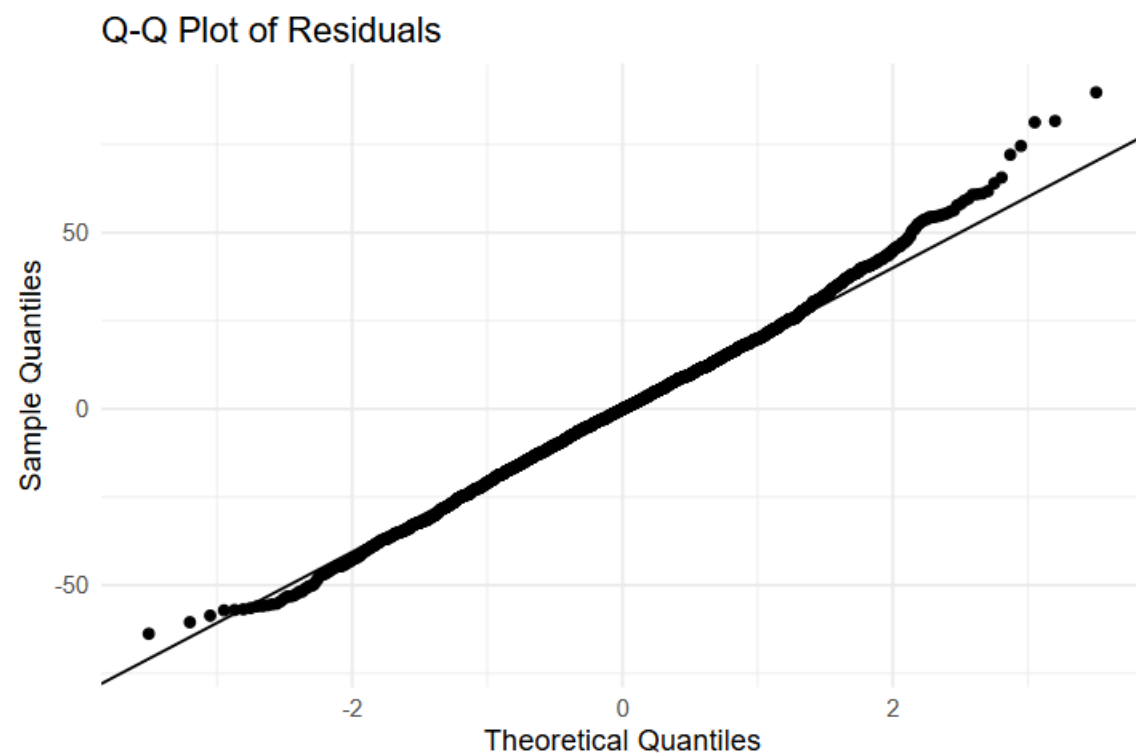
Appendix F Figure 14: Dosage analysis residual Q-Q plot



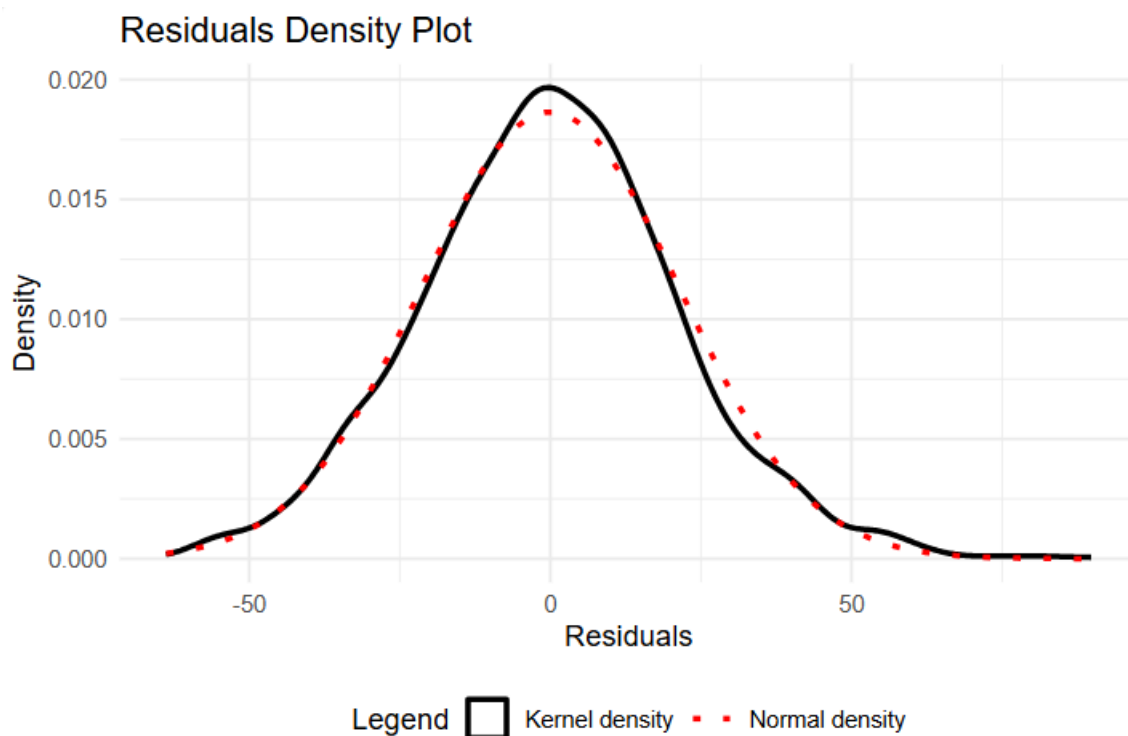
Appendix F Figure 15: Exploratory compliance analysis 1 residual density plot



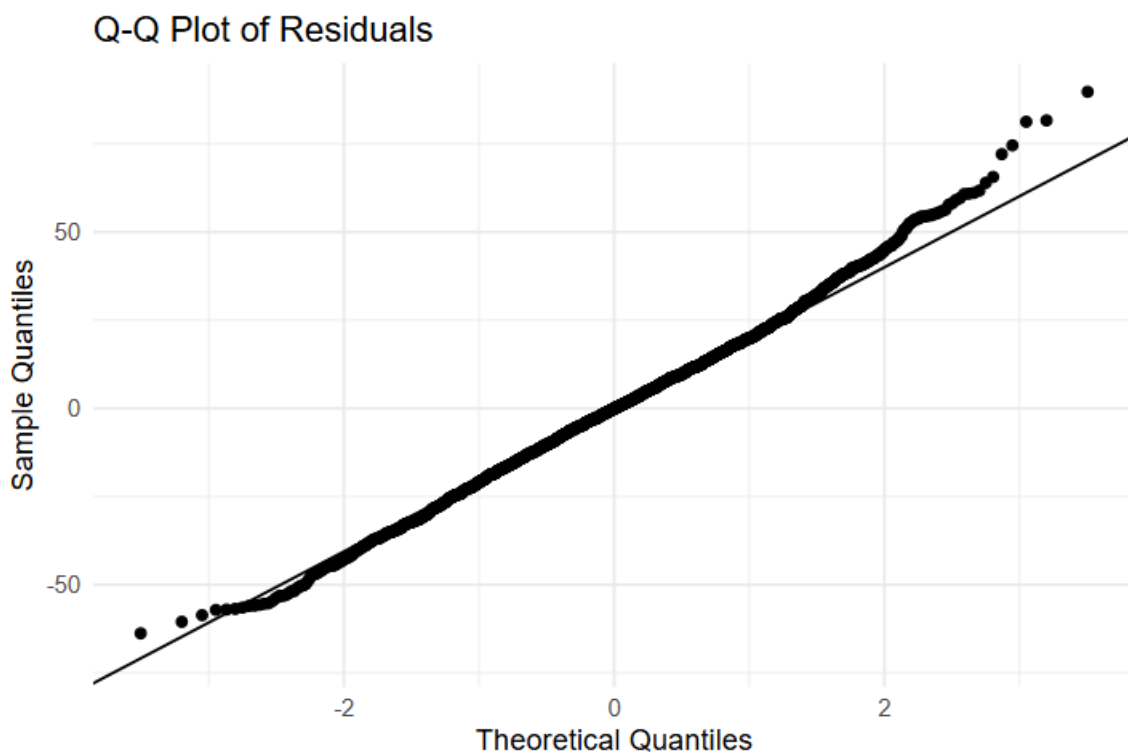
Appendix F Figure 16: Exploratory compliance analysis 1 residual Q-Q plot



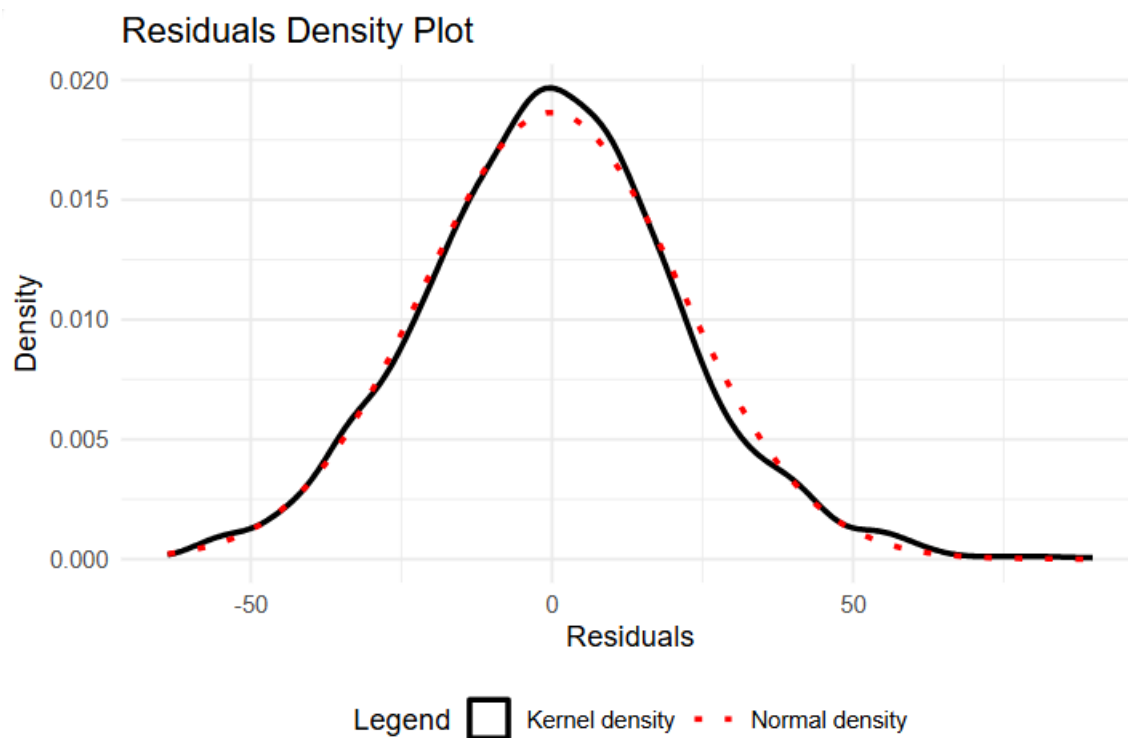
Appendix F Figure 17: Exploratory compliance analysis 2 residual density plot



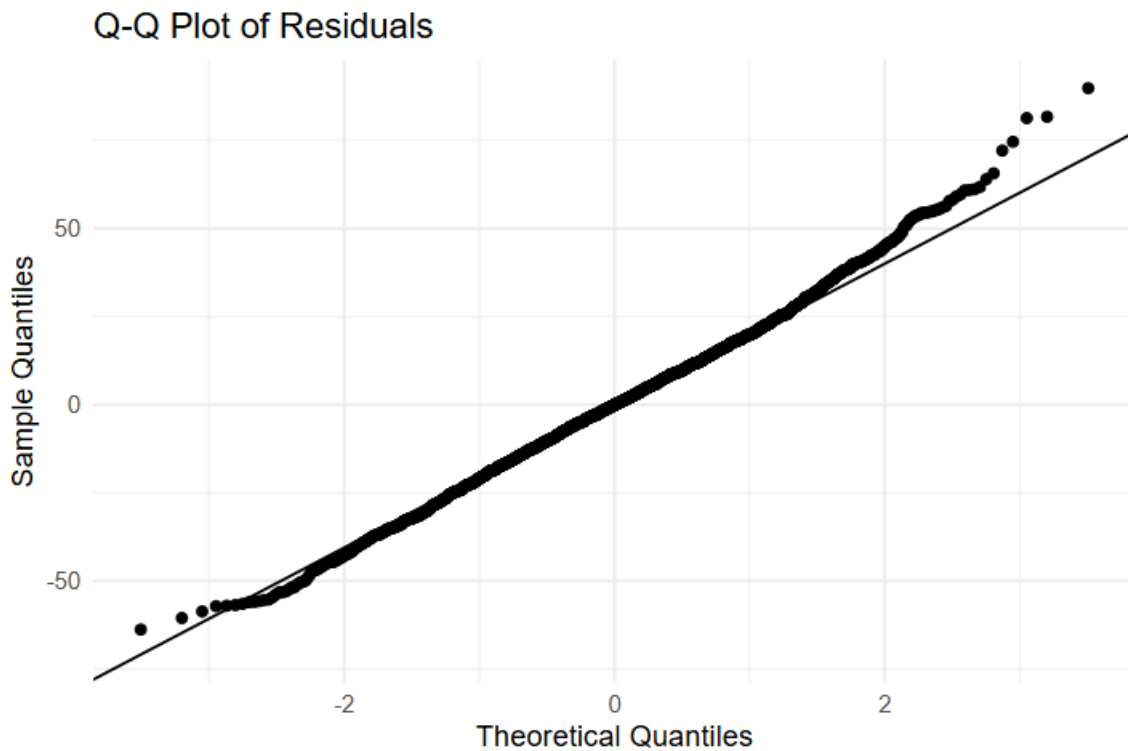
Appendix F Figure 18: Exploratory compliance analysis 2 residual Q-Q plot



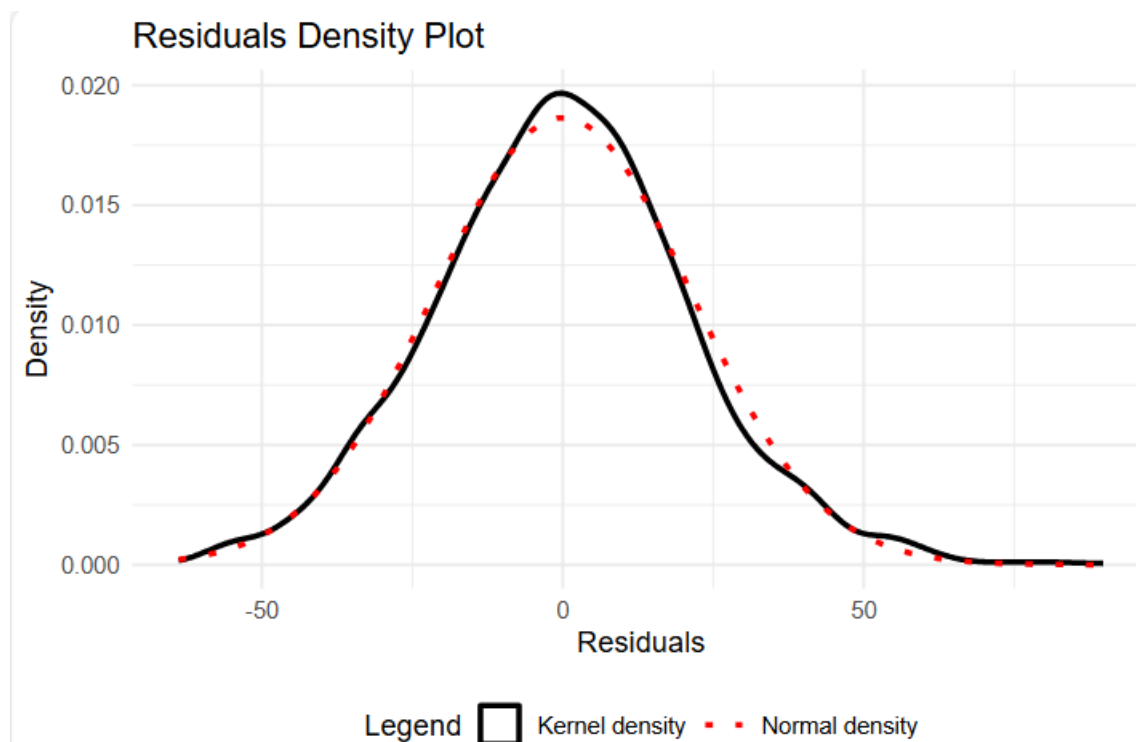
Appendix F Figure 19: Exploratory compliance analysis 3 residual density plot



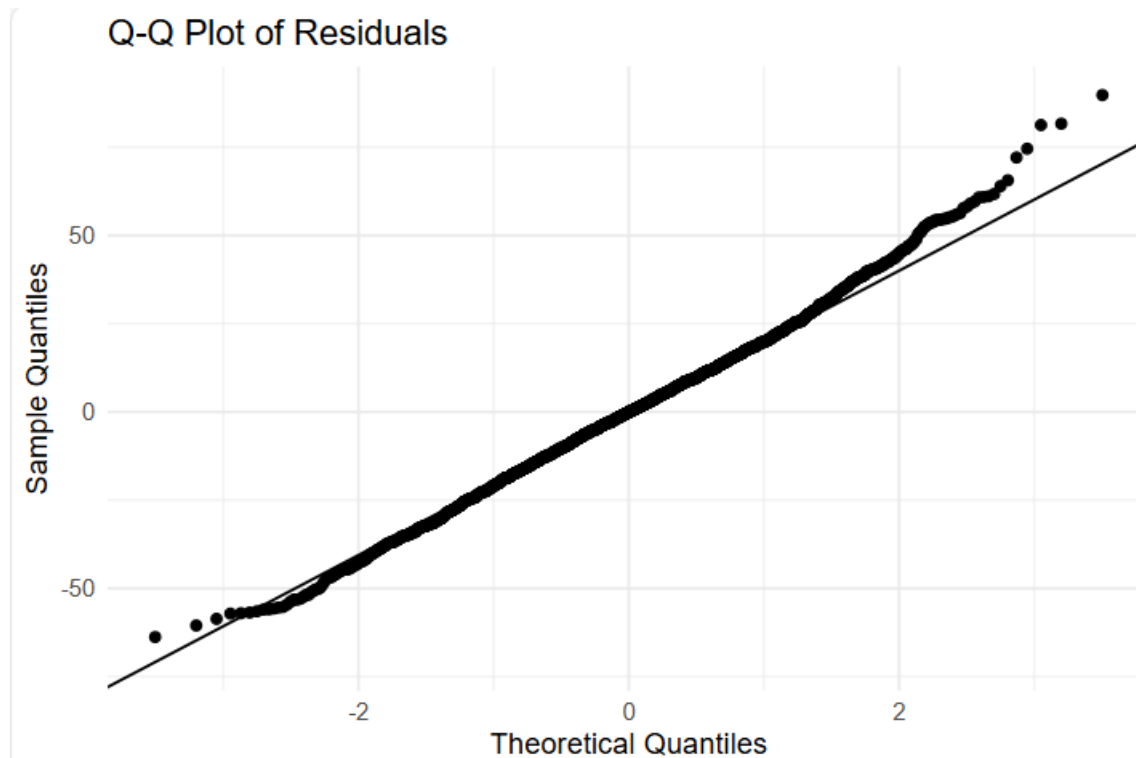
Appendix F Figure 20: Exploratory compliance analysis 3 residual Q-Q plot



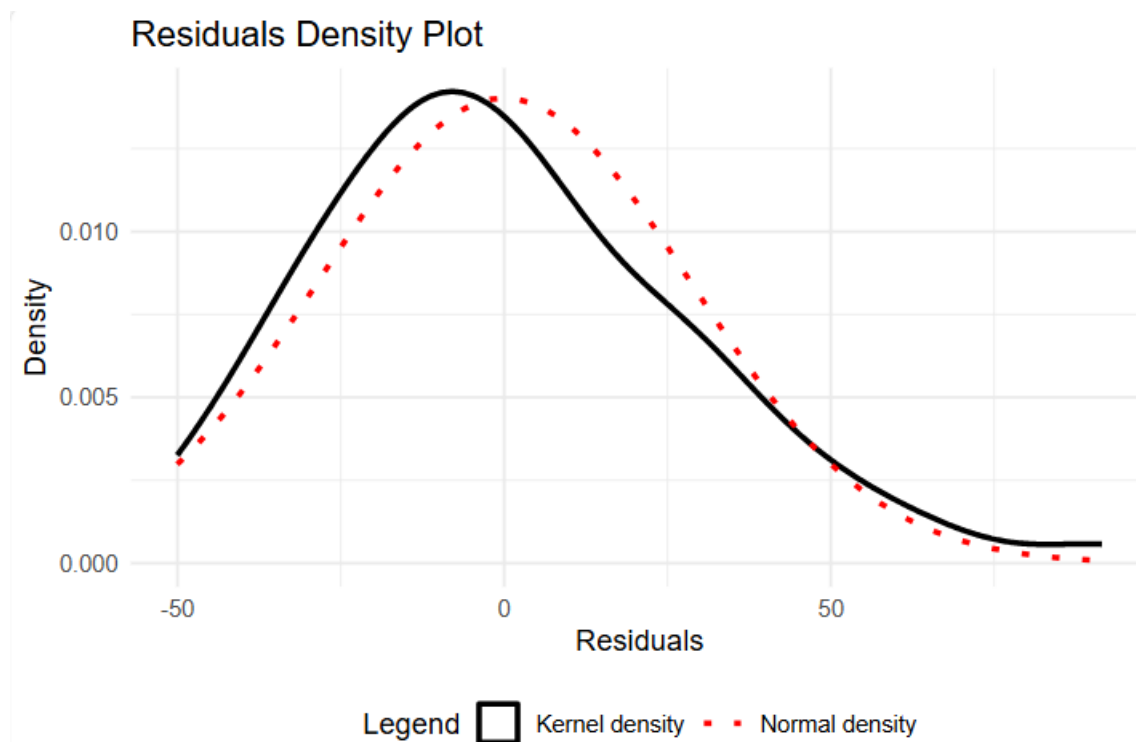
Appendix F Figure 21: Exploratory compliance analysis 4 residual density plot



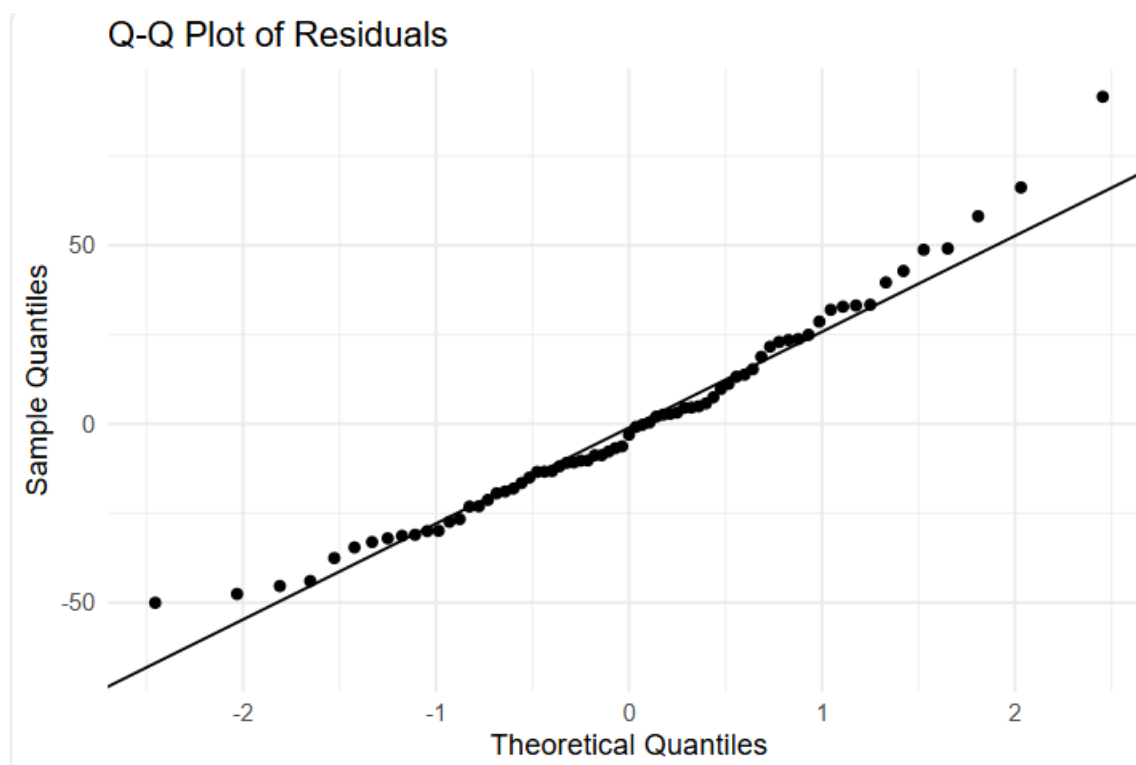
Appendix F Figure 22: Exploratory compliance analysis 4 residual Q-Q plot



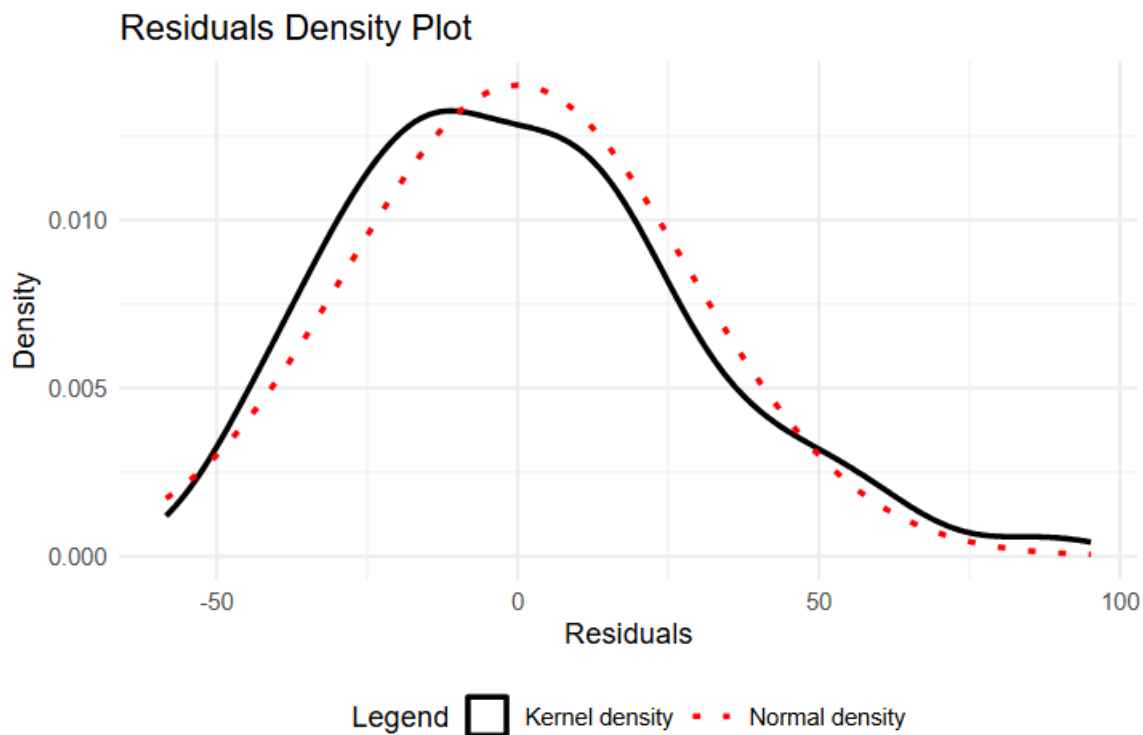
Appendix F Figure 23: Bottom 20% analysis residual density plot



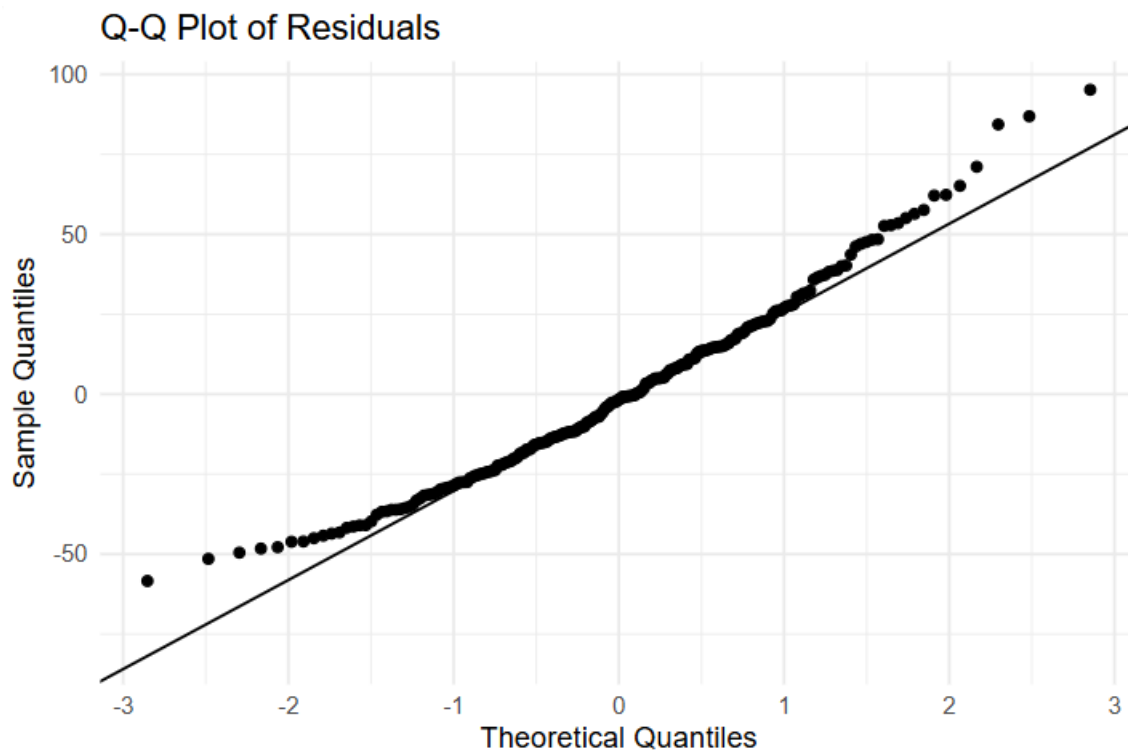
Appendix F Figure 24: Bottom 20% analysis residual Q-Q plot



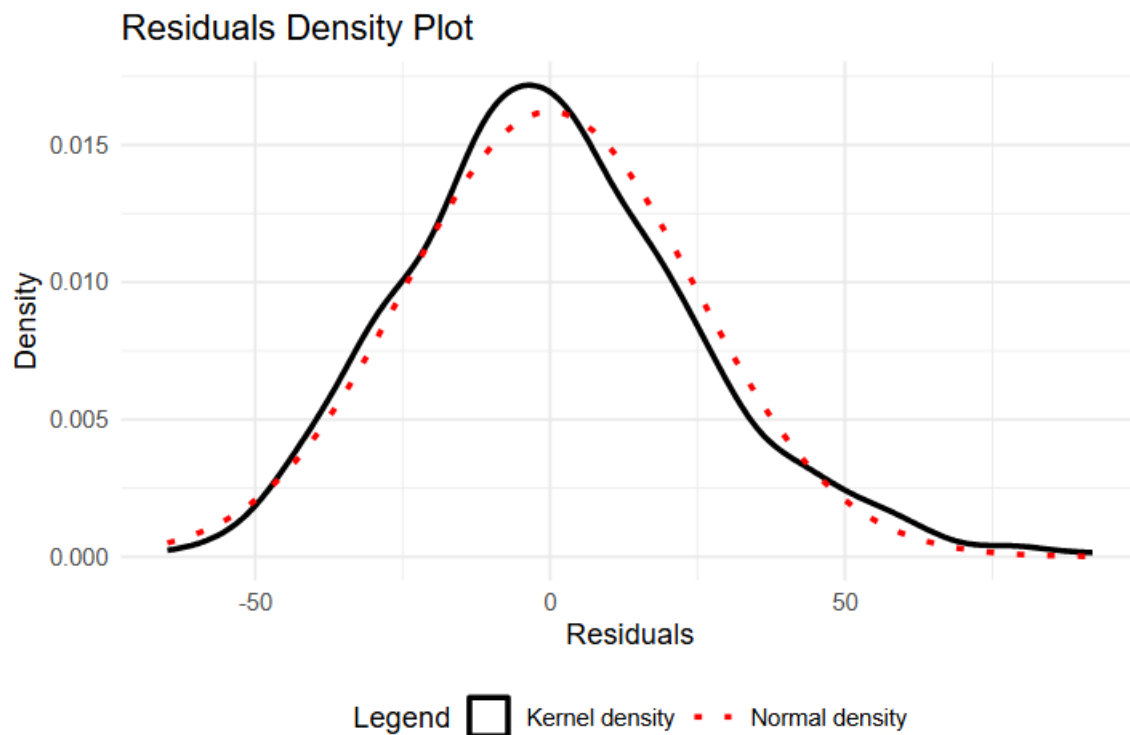
Appendix F Figure 25: Bottom 30% analysis residual density plot



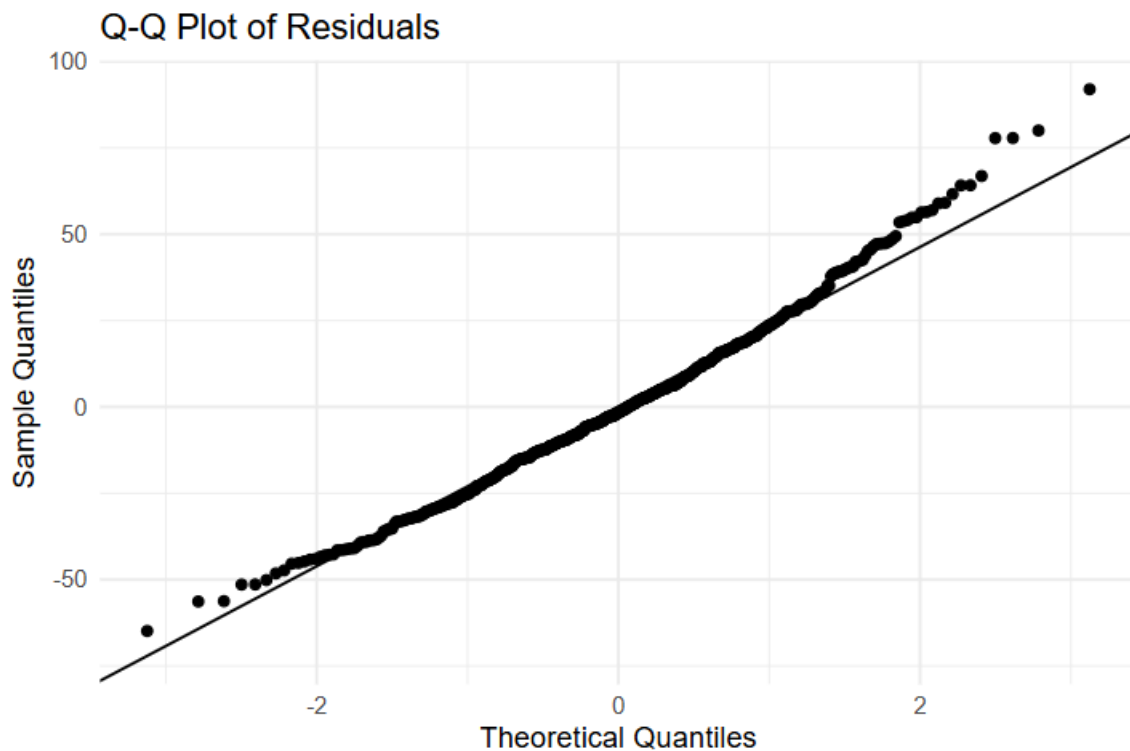
Appendix F Figure 26: Bottom 30% analysis residual Q-Q plot



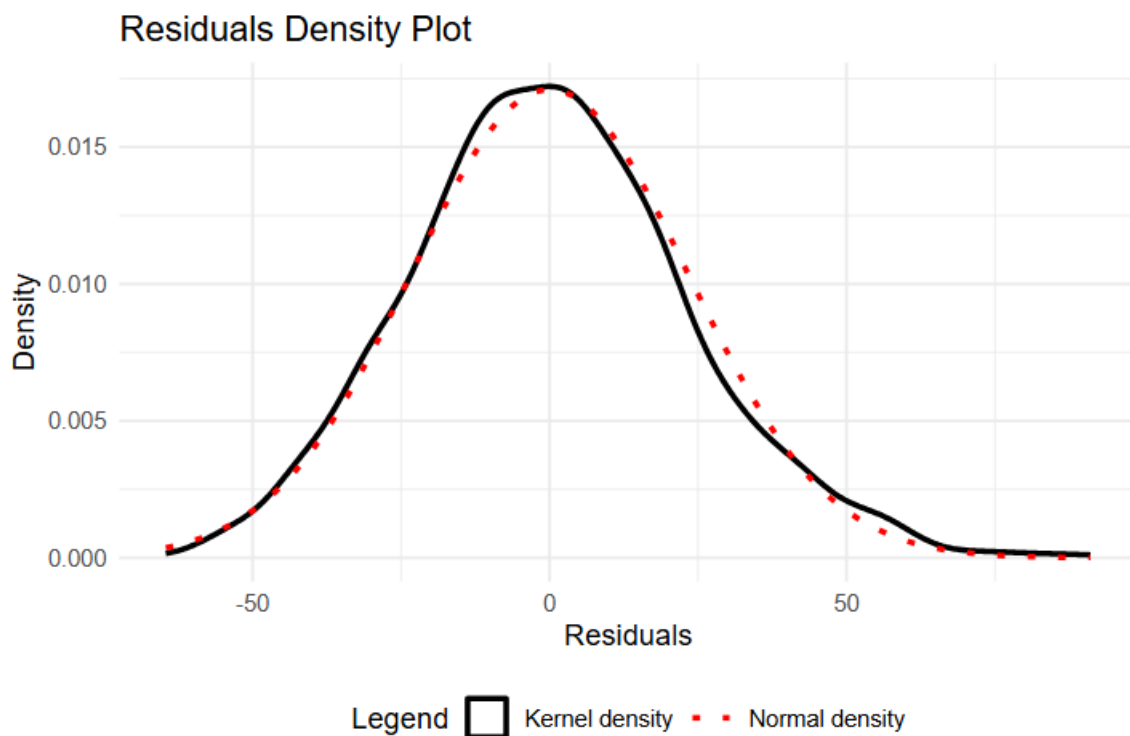
Appendix F Figure 27: Bottom 40% analysis residual density plot



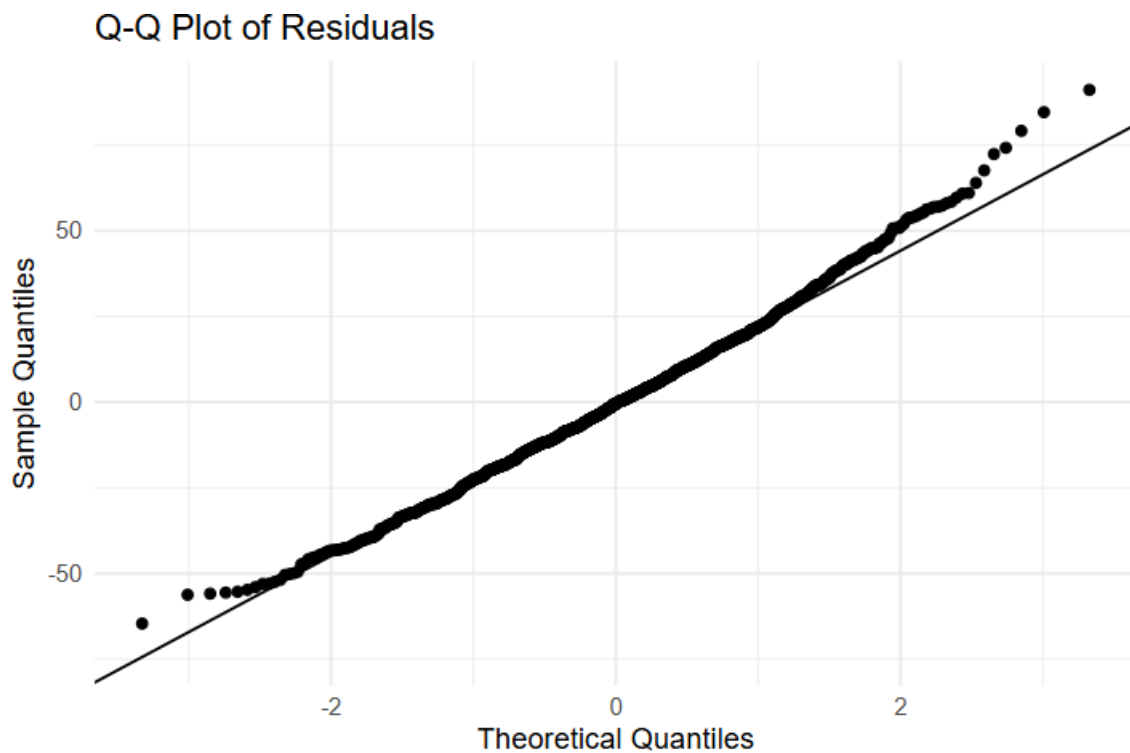
Appendix F Figure 28: Bottom 40% analysis residual Q-Q plot



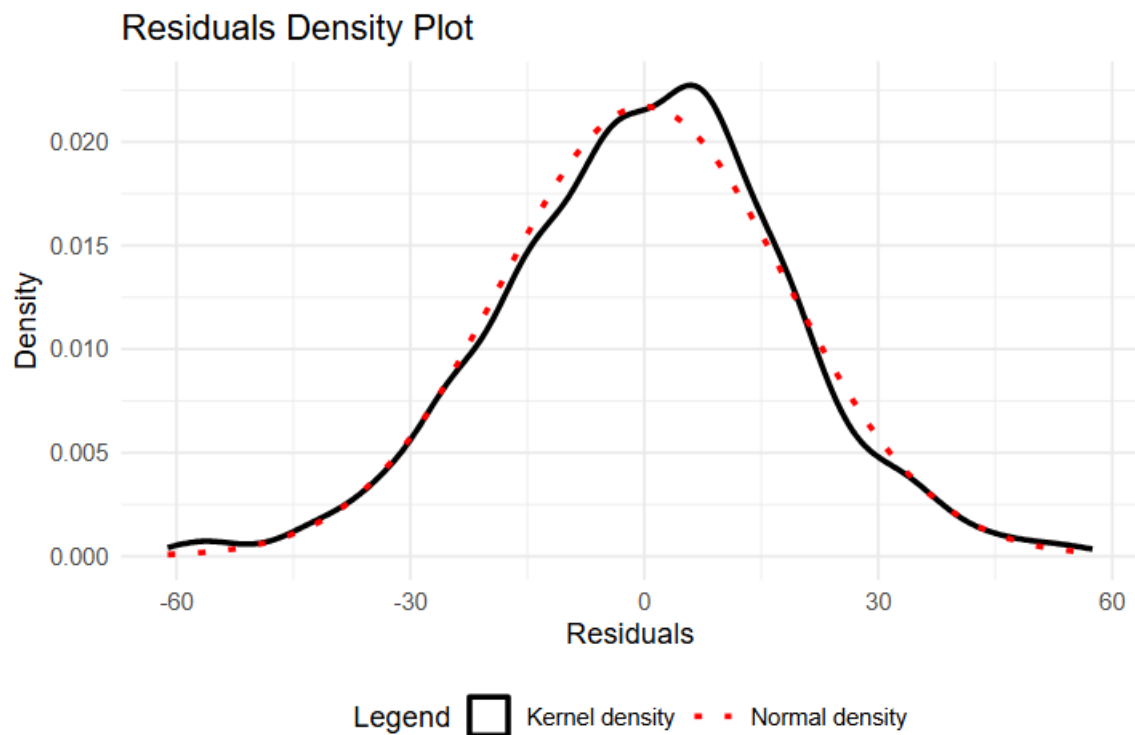
Appendix F Figure 29: Bottom 50% analysis residual density plot



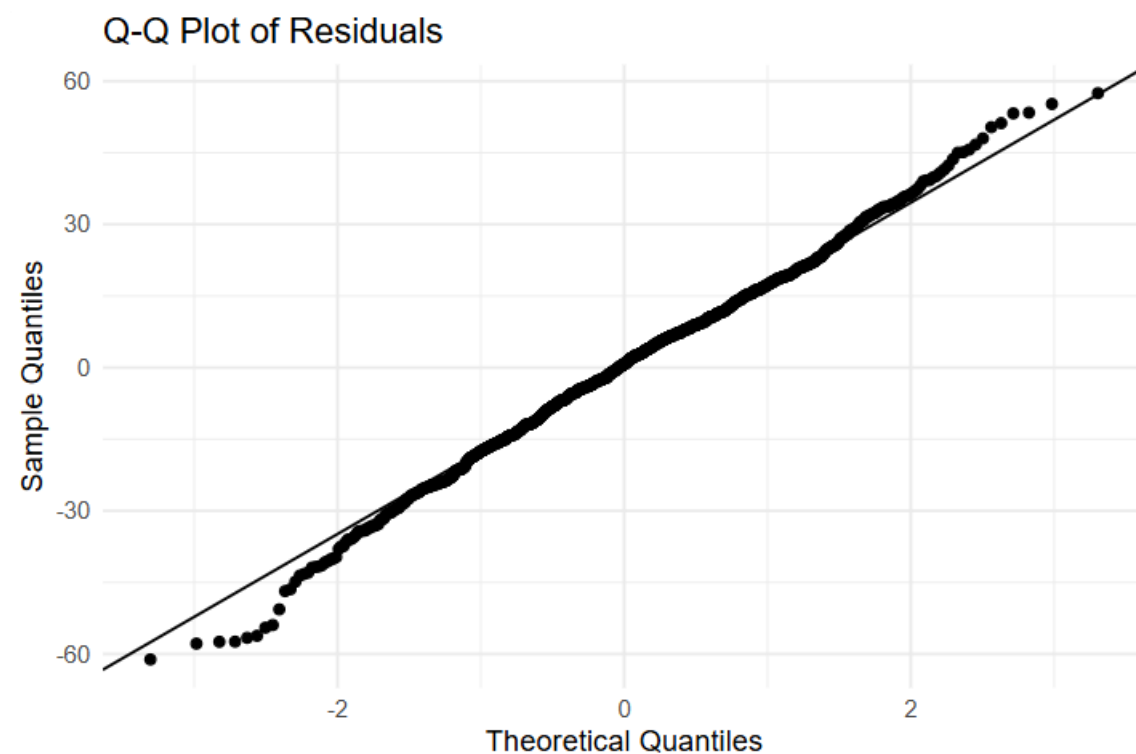
Appendix F Figure 30: Bottom 50% analysis residual Q-Q plot



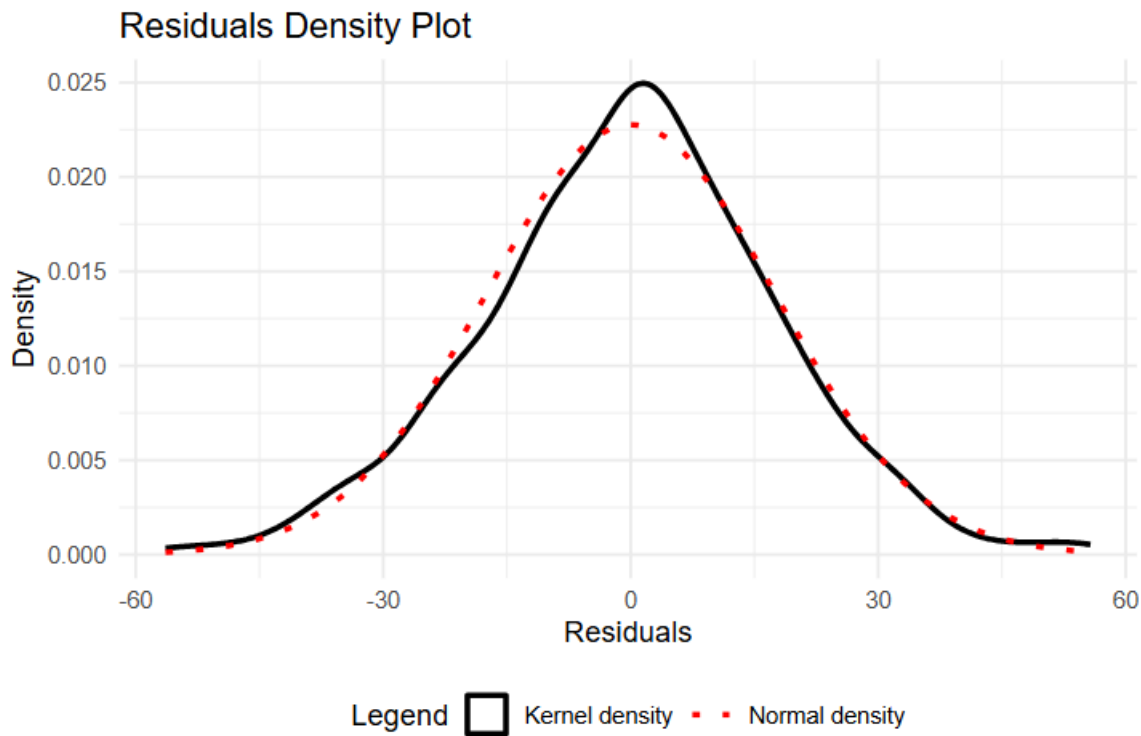
Appendix F Figure 31: Top 50% analysis residual density plot



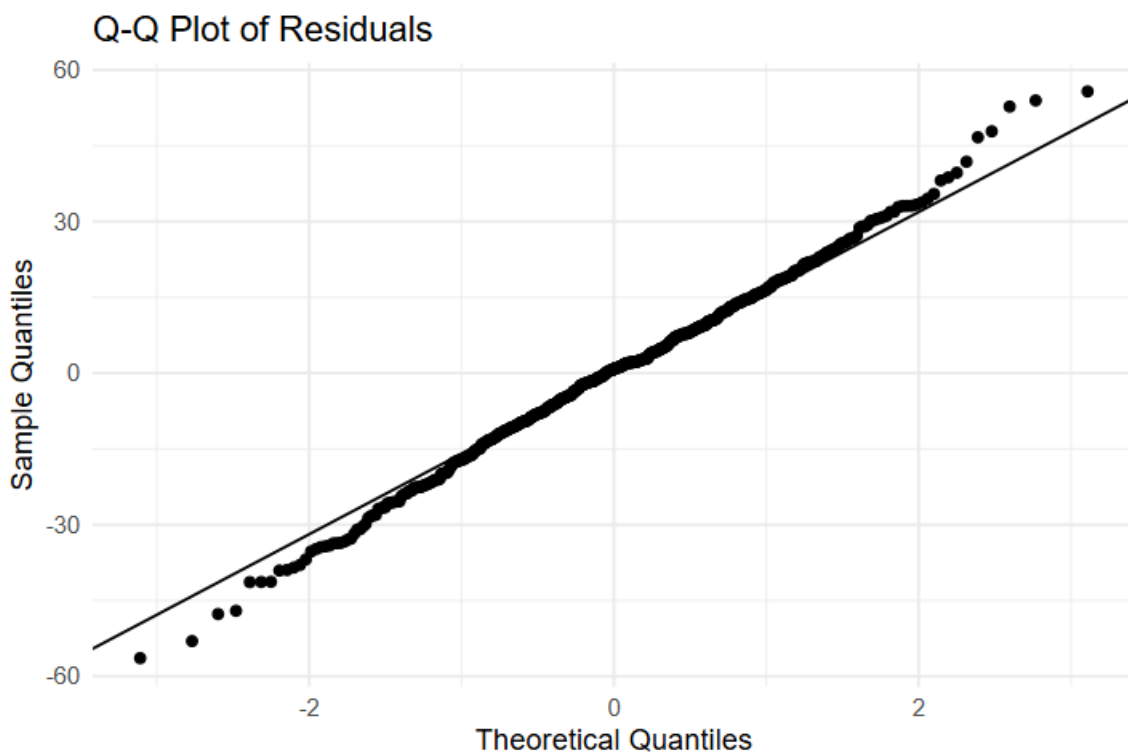
Appendix F Figure 32: Top 50% analysis residual Q-Q plot



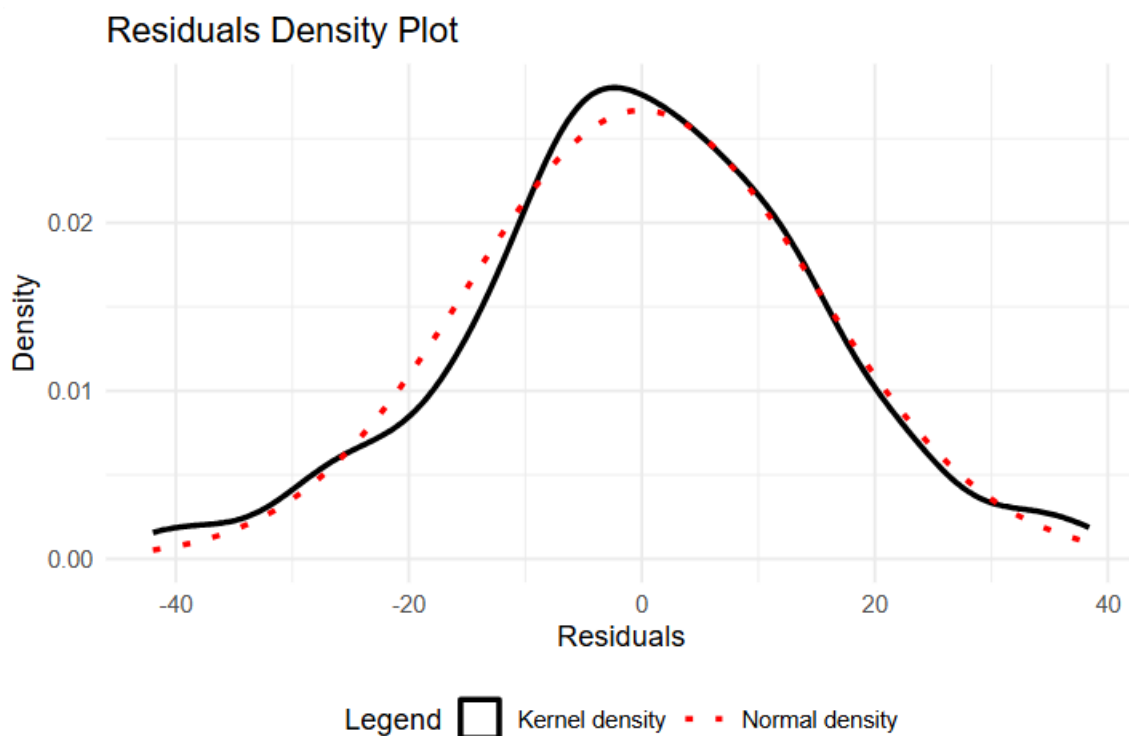
Appendix F Figure 33: Top 40% analysis residual density plot



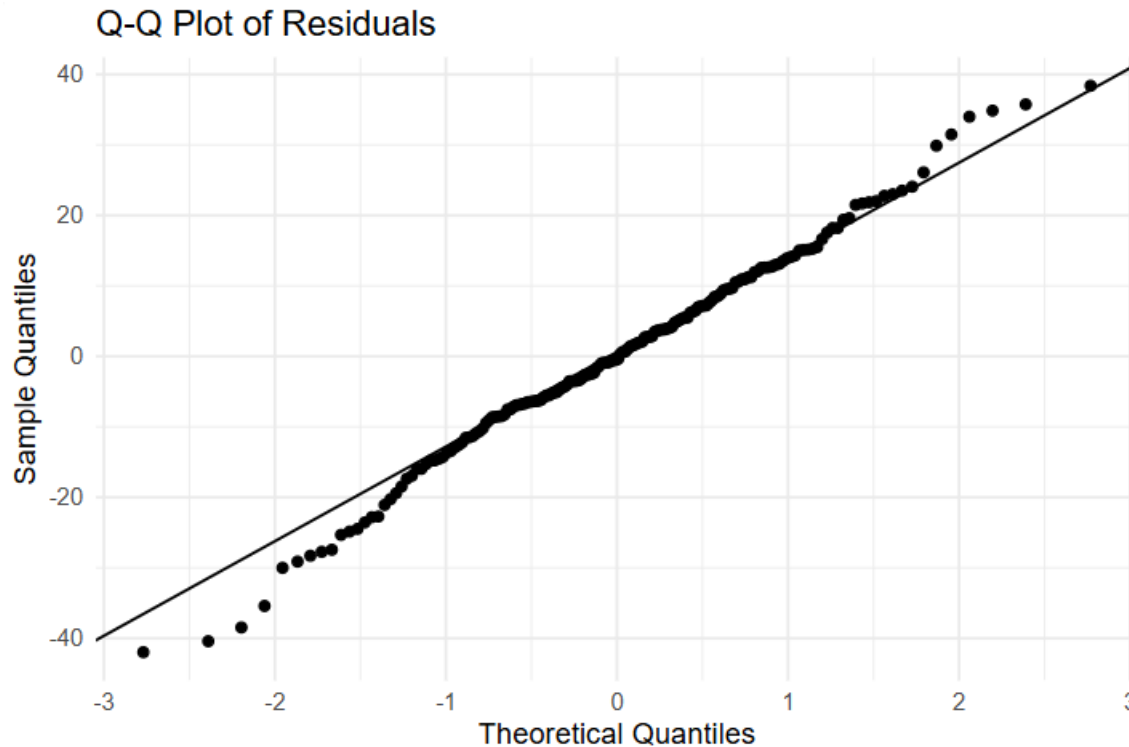
Appendix F Figure 34: Top 40% analysis residual Q-Q plot



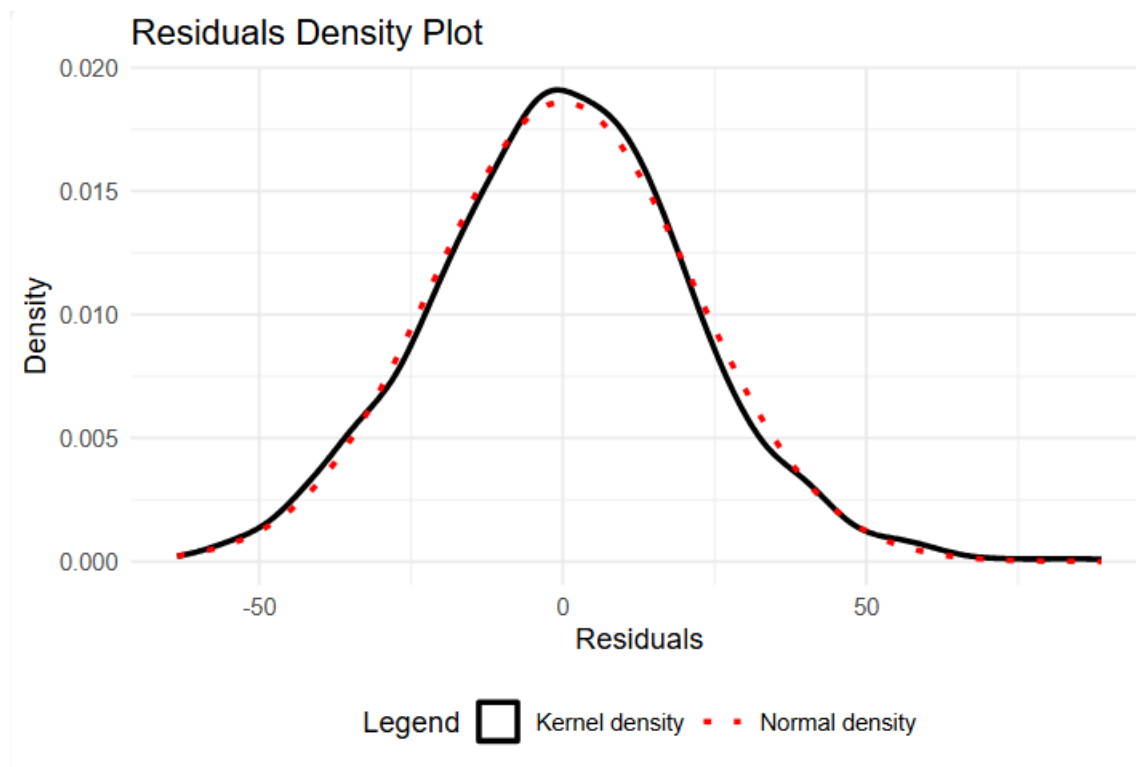
Appendix F Figure 35: Top 30% analysis residual density plot



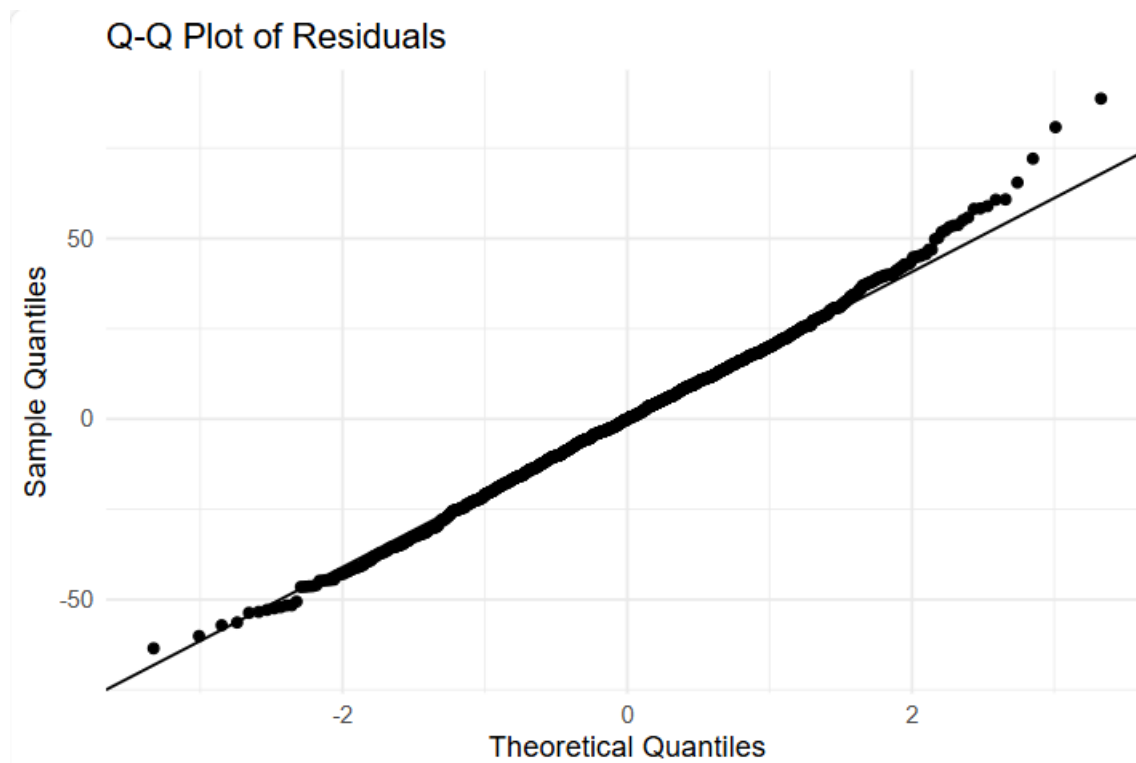
Appendix F Figure 36: Top 30% analysis residual Q-Q plot



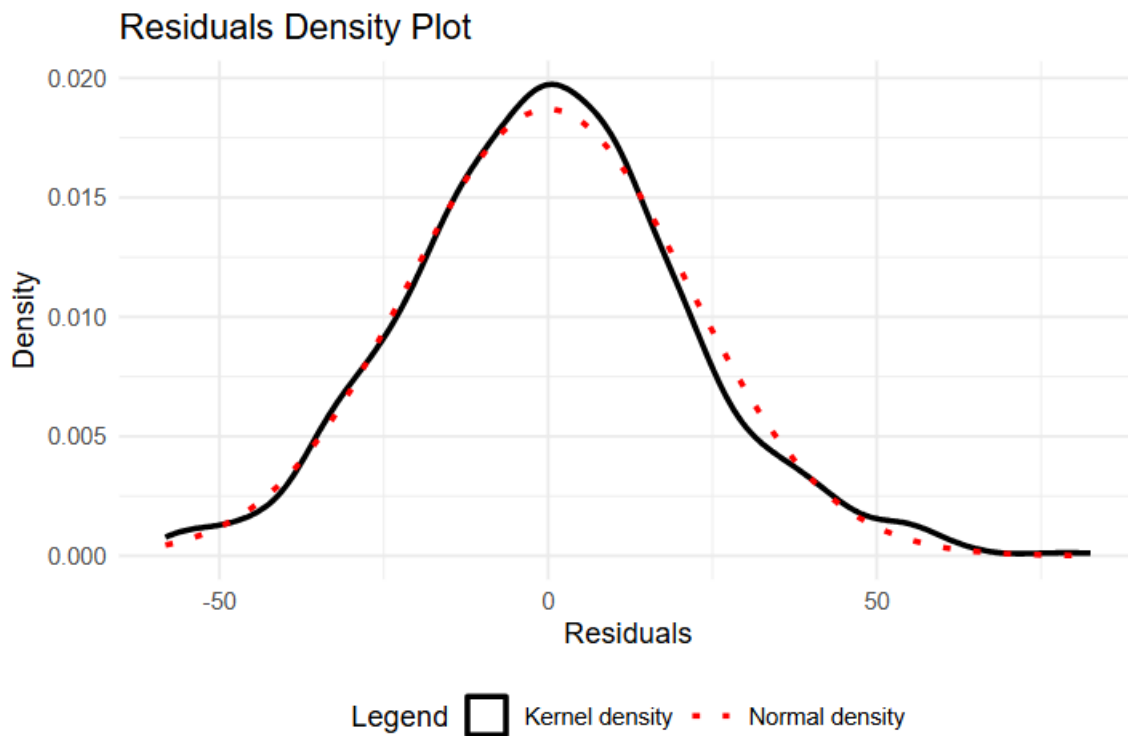
Appendix F Figure 37: Cohort 1 analysis residual density plot



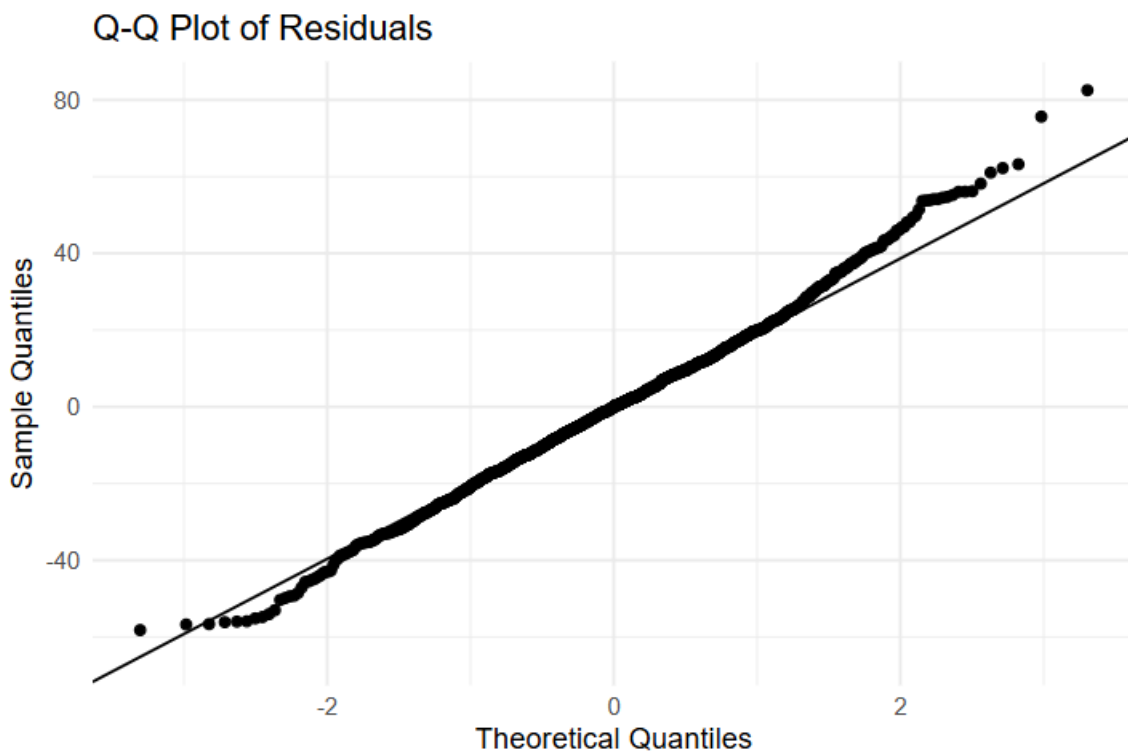
Appendix F Figure 38: Cohort 1 analysis residual Q-Q plot



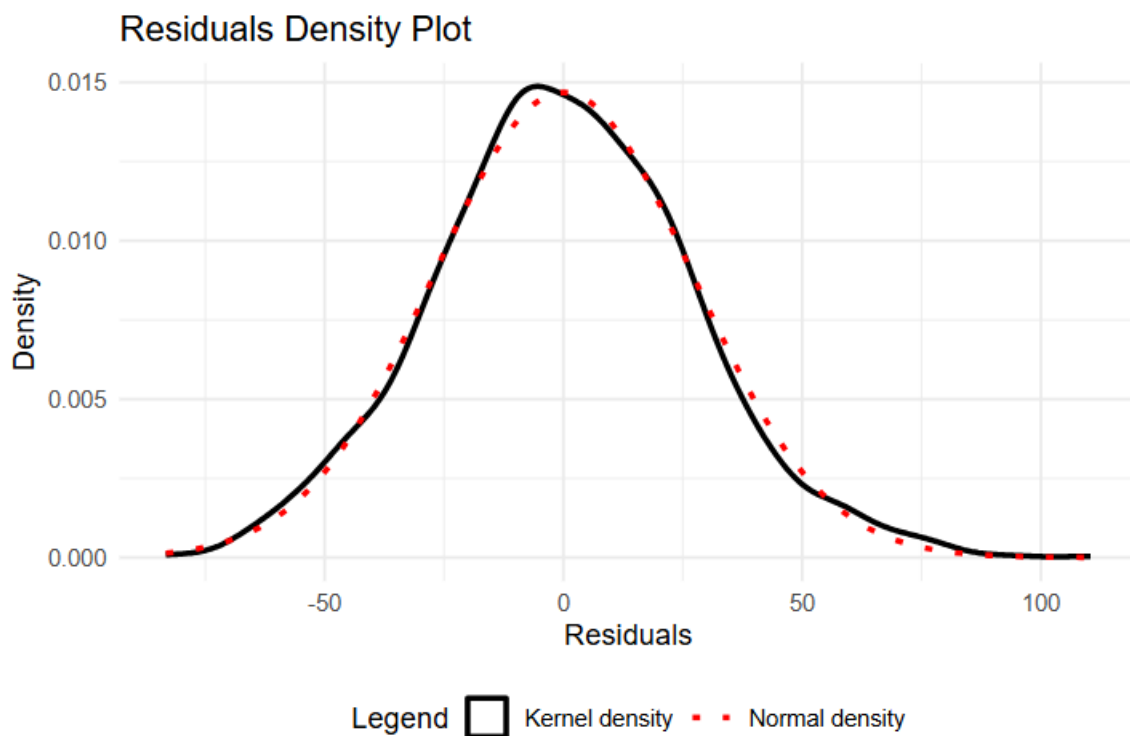
Appendix F Figure 39: Cohort 2 analysis residual density plot



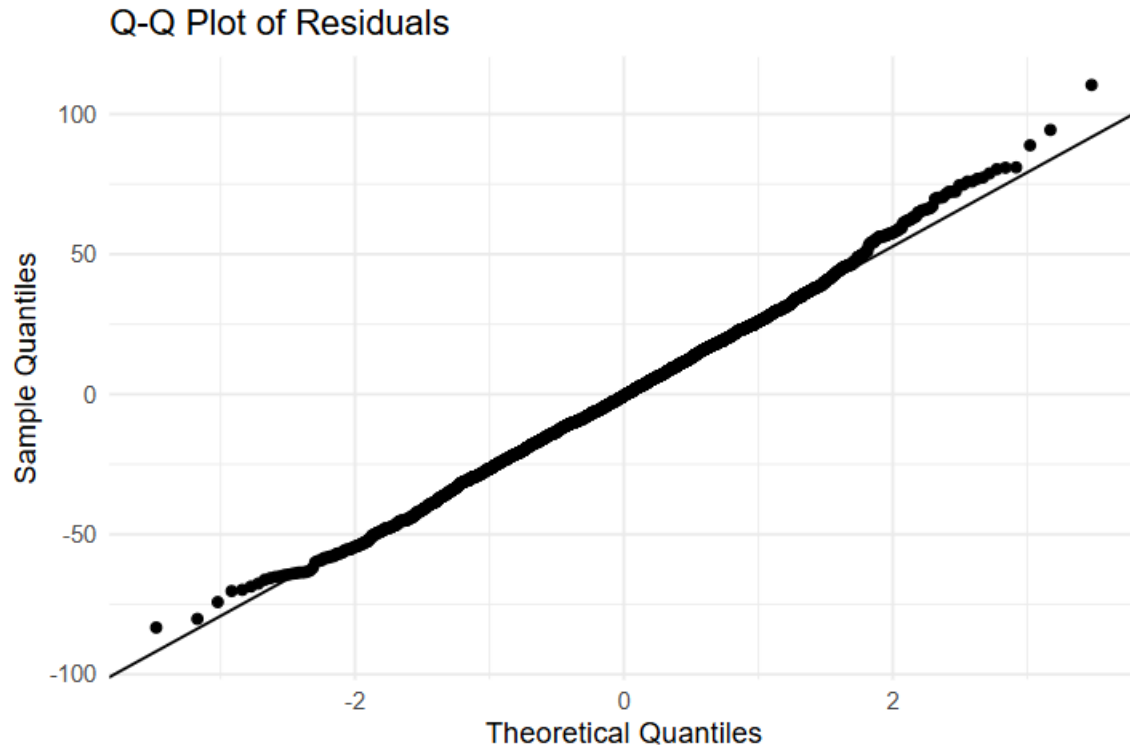
Appendix F Figure 40: Cohort 2 analysis residual Q-Q plot



Appendix F Figure 41: Oral Reading Fluency analysis residual density plot



Appendix F Figure 42: Oral Reading Fluency analysis residual Q-Q plot



Appendix G: Memorandum of Understanding

LEXIA® CORE5® Reading Randomised Controlled Trial Study

MEMORANDUM OF UNDERSTANDING [Cohort 1]

This memorandum of understanding (MoU) explains what your school's participation in this research trial will entail. If you agree to take part and accept the terms and conditions outlined below, please sign a copy of this form and return by email to lexia@qub.ac.uk

This project is part of a wider DfE funded programme called the 'Accelerator Fund' (AF). The AF aims to increase access to evidence-informed programmes.

The details of the member of staff signing the MoU (name and work email) will be shared with the Education Endowment Foundation (EEF). Please see the [EEF's privacy notice](#) for more information on how EEF processes and uses your data.

The teams

During this project, you will be contacted by both the **Project team** (Queen's University Belfast and LexiaUK) for the delivery of the programme, and by the independent **Evaluation team** (RAND Europe and Elklan), who are carrying out an evaluation of its effectiveness.

Project team

Queen's University Belfast (QUB) lead school recruitment and are working in partnership with LexiaUK, who will provide software licences and training for schools allocated to the intervention group (see more detail on project team roles and responsibilities on p.7).

Evaluation team

RAND Europe are leading the independent evaluation of Lexia® CORE5® Reading and working in partnership with Elklan who will administer the testing at the beginning and end of programme delivery (see more detail on evaluation team roles and responsibilities on p.7).

What is the Lexia® CORE5® Reading programme?

[Lexia® CORE5® Reading](#) (Lexia) is a computer-based Independent Learning System (ILS) for reading, provided by LexiaUK. Lexia provides personalised practice and instruction in all key skill areas which make a successful reader (R-Y6). As pupils work independently, Lexia tracks progress and adapts automatically to provide extra practice and instruction. In addition, if required, the system suggests teacher-led practice and instruction through its bank of printable resources.

A previous [EEF-funded evaluation of Lexia](#) found that struggling readers made on average one additional month progress in reading compared to struggling readers who did not receive Lexia. Based on these findings, the EEF have funded this second, larger research project to test the programme's effectiveness when delivered at greater scale.

What training do school staff receive?

Lexia is an Independent Learning System, designed to provide personalised learning to each of the Year 2 pupils selected. School staff teams including teacher lead and teaching assistant delivery staff will be trained to set up intervention group pupils on the Lexia system, to monitor and provide instruction as required, and to understand the data provided by the programme.

Senior leader(s) will also be trained on implementation best practice. LexiaUK will provide all training via online conference facilities. A senior leader, identified lead Lexia teacher, and teaching assistants working with the intervention group will attend one introductory session (1 hour) and up to a further three training sessions over the course of the intervention period (allow up to 1hr 30 min per training session, for typically 6 staff per session).

Once the Lexia licence is issued to your school and Lexia is set up on your system, school staff will receive online training from LexiaUK. Trained school staff will start delivering Lexia to the selected Year 2 pupils immediately following training.

What happens during delivery?

During delivery, the school is expected to schedule **4-5 sessions of 30 minutes (including 10-minutes log-on/log-off time) per week for the delivery of this targeted intervention over 24 weeks with the selected Year 2 pupils (intervention group)**. The intervention group will be withdrawn from their class and supervised by a trained member of staff.

Staff will log on to reports at least once per week in order to check progress, and if necessary, deliver the printable resources which the system provides. Senior leadership should log on once per week to check fidelity of implementation.

What technology is required?

Lexia can be accessed on a web browser and iPad. A reliable internet connection is required. Pupils should also use headphones and have a quiet area to work in during sessions. More information on IT system requirements are available [here](#).

Although in usual practice, home use is available, this is not permitted until after the 24-week intervention period.

Why is this being evaluated?

Building on previous evidence of impact, the EEF are funding a new randomised controlled trial of Lexia to test the programme's effectiveness when delivered at greater scale.

Although some schools use Lexia across the age and ability range, it is more commonly used as a targeted intervention for pupils, which is how we will evaluate the programme in this research trial, with struggling readers from Year 2, selected by their teachers.

This evaluation will seek to measure the effectiveness of Lexia in improving the reading outcomes of these Year 2 pupils. In doing so, it will compare the impact of Lexia on pupils' outcomes against usual reading instruction. Furthermore, the evaluation will also probe into contextual factors that influence the delivery of Lexia across different types of schools. Since this evaluation follows from an efficacy study, it will build on the findings of the former.

Why a randomised controlled trial?

The EEF regularly evaluates its projects through an established evaluation process called Randomised Controlled Trials (RCTs). In an RCT, participation is determined by random selection. For this project, that means each school's involvement in the Lexia programme will depend on whether they are randomly

assigned into one of the two groups; the intervention group (who will deliver Lexia) or the control group (who will continue with their usual teaching).

What will happen in the trial?

The trial will include 224 primary schools in total, of whom half will be randomly assigned to deliver Lexia (intervention group) and half will be randomly assigned to continue with business as usual (control group).

The trial will take place over the course of two years:

Cohort 1: June 2023 to June 2024 (112 schools, 56 intervention group and 56 control group)

Cohort 2: June 2024 to June 2025 (112 schools, 56 intervention group and 56 control group)

The evaluation team will randomly allocate every school in the trial into one of two groups:

Intervention group

Schools allocated to the intervention group receive the Lexia programme.

All intervention schools will receive a Lexia licence (three-year term) from LexiaUK which includes the following:

- Independent/personalised online activities for pupils.
- Reporting tools to inform planning and instruction.
- Printable offline paper-based resources and teacher guides.
- Initial online training and ongoing online or phone support for school leaders and their staff who are involved in project delivery.

Intervention schools pay £350 including VAT to LexiaUK (to be paid before licence issue) to participate in the study and will be issued with a three-year licence for up to 50 pupils. This represents a small partial contribution to the full cost of this licence which stands at £4,620 (including VAT).

Control group

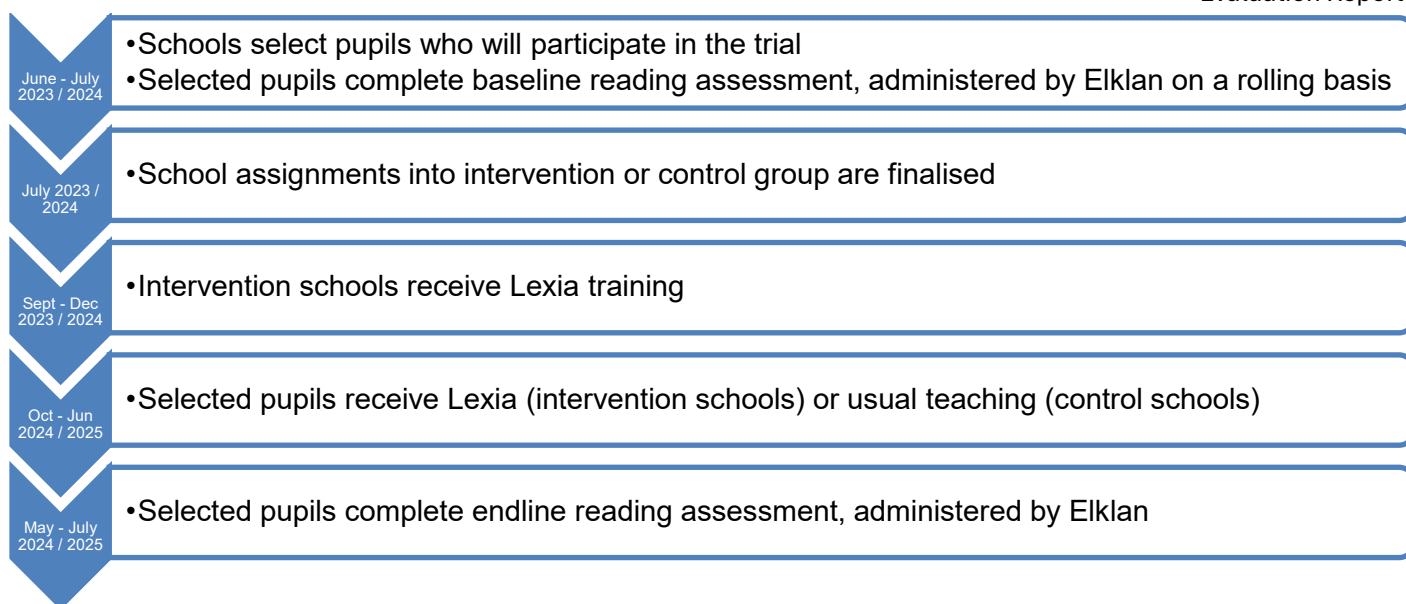
Schools randomly allocated to the control group do not receive the programme but make an important contribution to the research trial. The impact of the programme will be evaluated by comparing schools receiving the programme with the “business as usual” approach to reading in Year 2 in control group schools.

Please note that schools who are randomly allocated to the control group are not waitlisted to automatically receive the programme in cohort 2. The randomisation process takes place for each cohort.

- All control group schools receive a payment of £1,000 for their full participation in the trial evaluation requirements. This will be paid to schools at the end of the project in Autumn term 2024 (Cohort 1) or Autumn term 2025 (Cohort 2).
- Control schools refrain from purchasing and using the Lexia software while they are part of the research trial until July 2024 (Cohort 1) or July 2025 (Cohort 2).
- Schools will be free to use their payment towards the purchase of Lexia from July 2024 (Cohort 1) or July 2025 (Cohort 2).

When is the trial taking place and how many schools will be involved?

We will require you to select a group of struggling readers from your Year 1 class(es) to participate in Year 2. The trial will proceed along the following timeline:



In total, 112 schools will be recruited to participate as cohort 1 (September 2023-July 2024 delivery). They will be asked to arrange testing in June/July 2023 to be eligible for the study. Of these 112 schools, 56 will be randomly selected to deliver the intervention in their school, and 56 will act as control group and continue with usual teaching. All schools in this group will be asked to arrange testing in June/July to comply with the requirements of the evaluation.

How many Year 2 pupils in each school will participate?

All participating schools who sign up to participate in this trial will use LexiaUK's guidance to select a minimum of 8 pupils (one class entry schools) or 12 pupils (two class entry schools) and a maximum of 16 pupils in total per school.

Pupils selected for the intervention are those who are able to work independently for 20-30 minutes at a computer with minimal need for one-to-one support, who are falling below age-related expectation in their reading.

Is my school eligible?

To be able to participate in the trial, a school must meet the following criteria:

- Have a minimum of 8 pupils (and a maximum of 16 pupils) that would be eligible for Lexia
- Not currently involved in any other EEF trial focusing on Year 2
- Have not held a Lexia licence in the 12 months prior to July 2022 (for cohort 1) and July 2023 (cohort 2)
- Able to provide the necessary IT technology required in order to support the intervention.

What will be required of schools as part of their participation in the trial?

Prior to randomisation

Schools will:

- ☐ Provide contact details (valid email addresses and telephone numbers) of a main contact person (Headteacher), the designated lead Lexia teacher and teaching assistants delivering the programme, and the IT technician to the Project team (see table at end of this document).

- ☐ Provide an information sheet and privacy notice to parents that describes the study, the data that will be collected, and that offers parents the opportunity of withdrawing their child out of the study (the school will collate any responses that indicate this wish as noted below). RAND Europe will provide schools with the information sheet, privacy notice, and other relevant documents that will be provided to parents. The school will inform RAND Europe of any responses arising.
- ☐ Sign and return a Data Sharing Agreement (DSA) that will be provided by RAND Europe.
- ☐ Provide RAND Europe and LexiaUK with: (1) names of teachers and teaching assistants who will be trained and deliver Lexia (intervention schools only), (2) pupil data (including UPNs, name, DOB, FSM Ever, SEND, EAL, etc) for those pupils selected to be part of the evaluation in intervention and control schools, and (3) the names of any pupils whose parents have requested they are withdrawn from the data collection in intervention and control schools.
- ☐ For intervention schools: Participating teachers and teaching assistants will complete the online training and support provided. The headteacher or other senior management member should attend training to assist in planning and implementation.
- ☐ Facilitate the administration of the standardised reading assessment delivered by RAND Europe's partner, Elklan, in June-July 2023 (Cohort 1) or June-July 2024 (Cohort 2) to pupils selected to receive Lexia in the following year.

During the evaluation

Participating staff from all schools (both intervention and control schools) will:

- ☐ Update details of selected pupils during the study and contact details during the course of the evaluation (if appropriate).
- ☐ Complete a baseline survey, which will be administered by RAND Europe to the headteacher, the designated lead Lexia teacher, and nominated teaching assistants in November 2023 (Cohort 1) or November 2024 (Cohort 2).
- ☐ Complete an endline survey, which will be administered by RAND Europe to the headteacher, the designated lead Lexia teacher, and nominated teaching assistants in Summer 2024 (Cohort 1) or Summer 2025 (Cohort 2).
- ☐ Facilitate the administration of the second standardised reading assessment delivered by RAND Europe's partner Elklan, this time in the summer term 2024 (Cohort 1) or summer term 2025 (Cohort 2) with the participating pupils in Year 2.
- ☐ If the school withdraws from the delivery of the project for operational or other unavoidable reasons, it will notify the Project Team straight away and wherever possible will continue with testing in the Summer term to provide data for the evaluation (2024 for cohort 1, summer 2025 for cohort 2).

Participating staff from intervention schools will:

- ☐ Contribute £350 (including VAT) for the Lexia licence, which is non-refundable once the licence is issued.
- ☐ Set up the Lexia programme software promptly and complete all online training and support sessions delivered by LexiaUK, keeping in regular contact when necessary.
- ☐ Give RAND Europe and LexiaUK permission to access the school's Lexia data through 'my Lexia' (online portal) in order that both can monitor the usage of Lexia during delivery. Note: With the exception of Head and Lead teacher, all school staff will be granted read only access to their Lexia data over the intervention period to minimise the potential of data corruption/deletion.
- ☐ At the earliest opportunity, notify the Project Team if there are support or operational issues which could prevent the effective use of the approach.
- ☐ Allow researchers from RAND Europe to observe a predetermined number of the training sessions delivered by LexiaUK.
- ☐ For a purposively selected sub-sample: facilitate two school visits by one or two researchers from RAND Europe. During the first visit, which will take place on a predetermined date between

September 2023 and October 2023 (Cohort 1) or September 2024 and October 2024 (Cohort 2) researchers from RAND Europe will observe existing instruction for struggling readers. During the second visit, which will take place during delivery of Lexia, researchers from RAND Europe will observe a Lexia session. The observations will be followed by discussions with select staff trained by Lexia, a member of the senior management team.

Use of Data

This research will be conducted in compliance with current data protection legislation. Details on how each team will manage your data and their roles are described further below.

Data collected by the project team

QUB will act as data controllers for all school staff data (i.e. names, contact details) collected as part of the recruitment phase. Once the recruitment phase is complete, QUB will share this data with data controllers LexiaUK (to set up licences and deliver training) and RAND Europe (for the evaluation).

For more information about how QUB look after your information, how to access your rights and who to contact if you have any queries or concerns about data protection please visit the [Queen's University Belfast website](#).

LexiaUK will act as data controllers for staff and pupil data to enable LexiaUK to deliver the intervention. Details of how personal data will be used for this and Lexia licencing purposes can be found in LexiaUK's [privacy policy](#).

Data removal

Any school who wishes to get in touch to have their data removed, please contact lexia@qub.ac.uk in the first instance during the recruitment and trial delivery phase, who will notify all relevant parties including the evaluation team to remove the data.

Once the trial has concluded, QUB and LexiaUK will delete data that they are holding for both cohorts (outlined above under 'data collected by the project team').

At the end of the trial, if you wish to remove your data, please contact RAND Europe [Elena Rosa Brown at erbrown@randeurope.org or Miguel Subosa at msubosa@randeurope.org].

Data collected by the evaluation team

RAND Europe will act as data controllers and will be the main point of contact for any matters relating to the protection of all personal data once the recruitment is completed and will make decisions about how and what personal data is used in the evaluation.

As part of the evaluation, RAND Europe will collect information from schools about all the children that take part in the trial including child outcomes assessed through standardised reading assessments (as specified above).

Elklan (the testing partner), will act as data processors throughout the evaluation period and will process the data in accordance with the ways and purposes set by RAND Europe. Before data collection takes place, RAND Europe will provide schools with a data sharing agreement (DSA) further outlining how data will be managed securely.

RAND Europe will provide schools with information sheets, privacy notices, and the option to withdraw from the research, for schools to share with the parents of potential participants and key school staff as

appropriate. These participant information sheets will be compliant with the requirements of the General Data Protection Regulation (GDPR), on the basis of legitimate interests for processing personal data.

Once DSAs have been signed, and information sheets, privacy notices, and withdrawal forms have been distributed to parents, schools will be invited to share the following information about children with the evaluation team: full name, date of birth, gender, unique pupil number (UPN). This data will be matched to participating children's scores on the Woodcock Reading Mastery Test, version 3 (WRMT-III), a standardised reading assessment, which will be administered by Elklan before randomisation and at the end of the evaluation.

What happens to the data when the trial ends?

At the end of the study, RAND Europe will submit the data to the Office for National Statistics Secure Research Service (ONS SRS) for archiving in the EEF data archive (managed by EEF's archive manager)¹² and will include data only individually identifiable to the Department for Education (DfE). At this point, the EEF will act as data controller for the archive. In future, this data may be matched to the National Pupil Database (NPD) and other administrative data during subsequent research.

Access to information will be restricted on a need-to-know basis and security arrangements are regularly reviewed to ensure their continued suitability. RAND Europe will collect and store all personal and special category data in accordance with the Data Protection Act (2018) and UK GDPR requirements. No personal information collected as part of this study will be transferred outside of the European Economic Area (EEA). Further details are provided in RAND Europe's Privacy Notice.

All results will be anonymised so that no school, staff member, or individual child will be identifiable in the report or dissemination of any results.

Responsibilities of the Project Team (Queen's University Belfast and LexiaUK):

- ☐ LexiaUK will provide the intervention schools with a three-year licence of Lexia® Core5® Reading
- ☐ LexiaUK will provide intervention schools with IT set-up guidance, and online staff training/support for delivery of Lexia with Year 2 pupils
- ☐ QUB will collect participating staff and lead contact names and email details and will share these with LexiaUK and the evaluator.
- ☐ QUB and LexiaUK will work closely with the Evaluation Team throughout the project

Responsibilities of the Evaluation Team (RAND Europe):

- ☐ Act as the first point of contact for any questions about the evaluation
- ☐ Conduct the random allocation of schools to the intervention and control groups
- ☐ Inform the schools whether they have been allocated to deliver the intervention or to act as a control group school.
- ☐ Provide information sheets, privacy notices, and withdrawal forms for the school to send to parents
- ☐ Collect class- and pupil-level data (including name, date of birth, UPN, Gender, WRMT-III baseline and endline scores)
- ☐ Conduct baseline and endline surveys with participating school staff
- ☐ Conduct interviews and focus group discussions with school staff from a selected subsample of participating schools
- ☐ Conduct two visits with a selected subsample of participating schools
- ☐ Request NPD data using pupil details

¹² You can find more information about the EEF archive on the EEF's website: <https://educationendowmentfoundation.org.uk/privacy-notices/privacy-notice-for-the-eeef-data-archive>.

- ☐ Analyse the data from the project and author the final report.

Please answer the following questions about your school

LA area and County	
School URN	
School admin email	
School address	
School telephone	
School Ofsted rating	
% FSM Ever pupils	
Number of classes in Year 2 (e.g. 1, 2, 3 etc)	
Are you currently using Lexia with any year group? If yes, please specify.	
Has the school been a Lexia customer in the past 2 years?	If Yes, Please specify when?

Key contact names and emails required		
Role during study	Contact name	Contact email
Head Teacher Assists in study implementation		
Main Contact (Head or Deputy) who will attend training and assist in study implementation.		
Lead Teacher who will act as Lexia Coordinator during the study		
IT Technician who will liaise with LexiaUK to ensure successful set up of online programme in school <i>Typically your outside IT agent</i>		
Teachers/HLTAs who will deliver the Lexia sessions to targeted Year 2 pupils		

HEADTEACHER AGREEMENT

Please initial each box and sign below:

- ☐ I confirm that I have read and understood the Memorandum of Understanding for the above project and evaluation and have had the opportunity to ask questions;
- ☐ I confirm I have read and understood training, implementation and delivery commitments.
- ☐ I agree to providing the data as specified in the Memorandum of Understanding and in the format requested by RAND Europe;
- ☐ I agree that this MoU and the data therein may be shared by QUB with LexiaUK and RAND Europe to facilitate the running of the project.
- ☐ I agree that data on students may be shared between the school, RAND Europe and LexiaUK to facilitate the running of the project
- ☐ I understand that failure to provide all the data specified as required prior to randomisation will prevent participation in the study. Any personal, identifiable data already provided will then be destroyed by RAND Europe.
- ☐ I agree to coordinate with RAND Europe's partner Elklan to arrange for two rounds of testing: Summer 2023 and 2024 (cohort 1) or Summer 2024 and 2025
- ☐ I agree to providing an information letter to all parents of all selected children and to inform RAND Europe of any parental opt-out from the study;
- ☐ I agree to sign the Data Sharing Agreement (DSA) and to provide RAND Europe with the relevant pupil details required for the study (i.e. pupil name, date of birth, unique pupil number)
- ☐ I agree to comply with the random allocation of the school to either act as control group or to be in the intervention delivery group;
- ☐ I understand that all data will be kept in accordance with the Data Protection Act (2018) and that no material which could identify individual children, teachers or the school will be used in any reports of this evaluation.
- ☐ If allocated to the intervention group, I agree to staff attending professional development online training, and to selected pupils fully participating in the Lexia programme.
- ☐ If allocated to the intervention group, I agree and understand that the contribution (£350 including VAT) paid by the school for the Lexia licence during this study is non-refundable once the licence is issued.
- ☐ If allocated to act as an intervention school, I understand that the LexiaUK licence is issued with the intention for the school to fully participate in the study for the duration of the trial between September 2023 and July 2024 (Cohort 1 schools) or September 2024 and July 2025 (Cohort 2 schools). The 3-year licence start date will be from the date of issue of Lexia licences.
- ☐ I confirm that the school does not currently hold a Lexia licence.
- ☐ I have read and understand LexiaUK's [Terms & Conditions](#).

I agree for my school _____ (Postcode _____) to
take part in the LexiaUK study and I accept the eligibility terms and conditions as described above
and agree to participate whether selected to act as an intervention school or a control school.

Signature of Head Teacher: _____

Name of Head Teacher: _____ Date: ____/____/____

PLEASE RETAIN A COPY FOR YOUR RECORDS AND RETURN A COPY BY EMAIL TO
Joanne O’Keeffe at lexia@gub.ac.uk, or contact this email address if you have any questions.
Please complete in electronic format.

Evaluation team contact details:

Elena Rosa Brown at erbrown@randeurope.org or Miguel Subosa at msubosa@randeurope.org

Appendix H: Parent information sheet, privacy notice, and withdrawal form

PARENT INFORMATION SHEET

Independent evaluation of the Lexia® Core5® Reading Programme

Your child's school has agreed to take part in a research study looking to examine the effectiveness of the Lexia® Core5® Reading Programme (hereafter called Lexia) in improving reading outcomes for struggling readers in Year 2.

This document tells you more about our research, so that you can understand why we are conducting it and what it will involve.

What is Lexia?

Lexia is a computer-based Independent Learning System (ILS) for primary school readers provided by LexiaUK. It provides personalised practice and instruction in all key skill areas which make a successful reader. Lexia includes practise and instruction in phonological awareness, phonics, fluency, vocabulary, comprehension skills, and spelling and grammar rules. As pupils progress through their personalised pathway, teachers can access progress reports which can inform planning and instruction. Furthermore, if the teacher feels additional support is required, a large bank of printable resources can be accessed, which includes lesson plans, independent worksheets, and class activity exercises.

Who is conducting this research?

The research project is funded by the Education Endowment Foundation (EEF). The intervention is led by LexiaUK and Queen's University Belfast (QUB) whilst the evaluation will be conducted by Elena Rosa Speciani and her team from RAND Europe. Elklan will be asked by RAND Europe to administer the one-to-one tests.

What does the research involve?

This study will aim to measure if Lexia helps improve the outcomes of Year 2 struggling readers across participating schools. To do this, we will compare the outcomes of pupils receiving Lexia against those of pupils receiving usual reading instruction. We will also look at different factors that may influence Lexia's effectiveness.

We will conduct the study with two cohorts, split across two academic years (2023/24 and 2024/25). In total, 224 schools across England will participate in the study, 112 of which will participate in 2023/24 and the remaining 112 in 2024/25. An independent team from RAND Europe will evaluate Lexia in the form of a randomised controlled trial (RCT). This means that half of the participating schools will be randomly assigned to administer Lexia whilst the other half will continue with usual reading instruction.

Children in schools assigned to administer Lexia will participate in the programme after selected teachers and teaching assistants (TAs) have been trained to use the software. Therefore, children's participation in the programme will start in September 2024 and end in June 2025.

What will happen if your child takes part in our research?

We plan to collect data from your child's school (name, gender, date of birth, eligibility for free school meals, unique pupil number). We will also ask that your child will complete a standardised child-friendly reading test, the Woodcock Reading Mastery Test (version 3; WRMT-III) twice: first between June and July 2024 and later between May and July 2025. These tests will be conducted one-to-one by test administrators from Elklan. Children will take the WRMT-III twice so that we can measure improvements before and after their participation in Lexia. Data we hold on participating children will be linked to the National Pupil Database (NPD), the government repository of educational data for all children in state-funded education in England. We will use the NPD to look at additional pupil information to probe further into whether Lexia has been effective and if our findings also apply across a range of different pupil characteristics.

In schools assigned to deliver Lexia, the children will participate in the programme from September 2024 to June 2025. Children in comparison schools will continue as usual. During programme delivery, members of the RAND Europe evaluation team may observe teachers and TAs in select schools whilst they implement Lexia. We will conduct these observations so we can monitor the quality of programme delivery, but we will not make any recordings. Your child may be participating in the intervention at the time, but no data will be collected on your child.

Please see the **Privacy Notice** at the end of this document for further detail on how personal data will be collected and used.

What happens to the results of the study?

All data for each child will be kept strictly confidential and only used for research, evaluation, and Lexia software usage purposes. Results from your child will be combined with the results from other children for statistical evaluation and will not be used in any way that would identify your child. At no time will your child's identity or individual data be reported. The team will not use your child's name or the name of the school in any report arising from the research, nor provide information that identifies your child to anyone outside the study team. The results of the research will be published on the EEF website.

Statement of any benefits or risk

Taking part in this research will be of no risk to your child. Since the pilot found the programme to have improved outcomes for struggling readers, children in schools assigned to deliver Lexia are expected to benefit from participating in the programme. The EEF is providing a grant to schools assigned to the comparison group, which schools can use at their discretion, including to purchase teaching programmes. The research will be of benefit to future classes of children.

What should I do next?

It is the school's decision to deliver Lexia, but it is your decision whether we can include your child in this research. Participation is entirely voluntary. Your decision will not adversely affect your child's access to services or programmes at your child's school outside the research. You are free to decline or stop at any time.

If you **DO NOT WISH** your child to take part in this study, please fill in the enclosed form and return it to your child's class teacher as soon as possible. If you are happy for your child to take part, there is no form to sign. If at any point, you decide to withdraw your child's data, you may do so without penalty and without giving a reason. To do this, you may contact the RAND Europe team, whose contact details are indicated on p. 8. These contact details will also be made available online. Upon notifying the RAND Europe team, we will delete your child's data and will not use it for the evaluation.

What if I have any questions or concerns?

If you have a concern about the delivery of Lexia, please contact Joanna O'Keefe at lexia@qub.ac.uk.

If you have any questions about testing and evaluation, please contact the evaluation team lead at RAND Europe, Elena Rosa Speciani, at lexia@randeurope.org.

If you have any questions on the way your child's personal data is processed, please contact RAND Europe's Data Protection Officer by emailing REdpo@randeurope.org.

We will acknowledge your concern within 10 working days and give you an indication of how it will be dealt with.

Privacy Notice for Parents (23035)

Independent Evaluation of the Lexia® Core5® Reading Programme

The delivery team (LexiaUK and QUB) and evaluation team (RAND Europe, who will employ and supervise the work of a testing company, Elklan, for the collection of outcome measures) conducting the Lexia study take privacy very seriously. Informed by the General Data Protection Regulation (GDPR), this privacy notice sets out how we will use and look after the personal information that we collect from the children and schools taking part in the study.

LexiaUK and RAND Europe are joint controllers of contact details collected from schools. Also acting as data controllers, QUB will be collecting these contact details during the study's recruitment phases. LexiaUK and RAND Europe also jointly control software use data and administrative pupil data collected for matching purposes.

RAND Europe is sole controller of:

- (a) more detailed demographic data on pupils, collected from both schools and the National Pupil Database;
- (b) data on the attendance of headteachers, teachers, and teaching assistants in compulsory training sessions conducted by LexiaUK; and
- (c) pupils' baseline and endline test scores on the Woodcock Reading Mastery Test, version 3, which will be administered by Elklan, who have been commissioned by RAND Europe as testing sub-contractors and data processors.

Once the study has concluded, data will be turned over to the EEF, who will then become the data controllers. However, the EEF are not data controllers until all data has been turned over to the EEF data archive.

Further data will come from the National Pupil Database (NPD). The NPD is the government's official repository of data on all children and young people in state-funded schools. The evaluation team will provide the Data Sharing Team at the DfE with the names of the pupils, their dates of birth, and Unique Pupil Numbers (UPNs), allowing a match to NPD and access to educational attainment and other relevant data. After the matching process has taken place, we will then analyse this data using the Secure Research Service (SRS) based at the Office for National Statistics (ONS). The SRS system does not allow users to remove or copy data from its servers.

The types of data collected, as well as the roles each party plays in the data processing conducted for programme delivery and evaluation are described in more detail in the table below.

All parties involved in data collection and analysis have put appropriate security measures in place to keep personal data secure and to prevent any unauthorised access to or use of it. This includes the use of password-protected folders for digital information and locked filing cabinets for written records.

What personal data will be collected, how, and what will it be used for?

Personal data means any information about an individual from which that individual can be identified. Data will be collected by both the delivery team and the evaluation team (with support from the testing company). The table below describes all the ways that personal data will be used by the delivery team and the evaluation team.

Personal Information Data				
Data (what will be collected?)	Source (who will collect?)	Receiver or Permitted recipient	Purpose (what will be used for?)	How do we collect the data?
Contact details (name, email address, telephone number) of headteacher, designated lead Lexia teacher, teaching assistants, and IT technician	QUB (data controllers)	LexiaUK and RAND Europe (data controllers)	LexiaUK: Transmission of documents relevant to the programme; setting up events related to the programme (e.g., training, software licence setup) RAND Europe: arranging school visits and data	Electronic data transfer via secure server

Personal Information Data				
Data (what will be collected?)	Source (who will collect?)	Receiver or Permitted recipient	Purpose (what will be used for?)	How do we collect the data?
			collection activities (e.g., survey)	
		Elklan (data processors)	Setting up dates for administering WRMT-III	
Basic administrative data of Year 2 pupils participating in the study, to be provided by schools (data already held by schools): • first name, last name • UPN • date of birth • gender	RAND Europe	LexiaUK and RAND Europe (data controllers) & Elklan (data processors)	Matching pupil records with software use data	Electronic data transfer via secure server
Additional demographic administrative data of Year 2 pupils participating in the study, to be provided by the schools (data already held by schools): • Free School Meal (FSM) eligibility status • Special educational needs and disabilities (SEND) status • English as an additional language (EAL) status	RAND Europe (data controllers)	RAND Europe (data controllers) & Elklan (data processors)	Matching pupil records; independent assessment of impact of intervention on outcomes; withdrawal forms	Electronic data transfer via secure server
		EEF (data controllers) & EEF archive manager (data processor)	Long-term follow-up analyses and archiving of the data. EEF becomes data controller once the data has been transferred to the EEF Data Archive but is not data controller until this has happened.	
Pupil data from the National Pupil Database: • FSM Ever6 status • SEND status • EAL status	National Pupil Database (data controller)	RAND Europe (data controllers)	Independent assessment (sub-group analyses) of impact of intervention on reading outcomes	Electronic data transfer via secure server
		EEF (data controller) & EEF archive manager (data processor)	Long-term follow-up analyses and archiving of the data. EEF becomes data controller once the data has been transferred to the EEF Data Archive	

Personal Information Data				
Data (what will be collected?)	Source (who will collect?)	Receiver or Permitted recipient	Purpose (what will be used for?)	How do we collect the data?
			but is not data controller until this has happened.	
Pupil outcome data collected for the evaluation (collected by test administrators): Woodcock Reading Mastery Test, version 3 (WRMT-III)	RAND Europe (Data controller) (Data collection sub-contracted to Elklan to test provider as Data Processor)	RAND Europe (data controllers) & Elklan (data processors)	Independent assessment of intervention impact on reading outcomes	Electronic data transfer via secure server
		EEF (data controller) & EEF archive manager (data processor)	Long-term follow-up analyses. EEF becomes data controller once the data has been transferred to the EEF Data Archive but is not data controller until this has happened.	
Lexia software use data	LexiaUK (data controllers)	RAND Europe (data controllers)	Analysis of compliance to programme design	Electronic data transfer via secure server
		LexiaUK	Matching pupil data with software use data	Electronic data transfer via secure server
		EEF (data controller) & EEF archive manager (data processor)	Long-term follow-up analyses. EEF becomes data controller once the data has been transferred to the EEF Data Archive but is not data controller until this has happened.	Electronic data transfer via secure server
Training attendance records (names of headteachers, teachers, and teaching assistants)	Schools (data processor)	RAND Europe (data controllers)	Analysis of compliance to prescribed training design and dosage	Electronic data transfer via secure server

Who will hold the data?

During the trial, the data will be managed by RAND Europe, LexiaUK, QUB, and Elklan. RAND Europe will securely transfer data to the EEF's archive at the end of the project under the EEF guidelines. At the end of the project, data will be submitted to the Office for National Statistics Secure Research Service (ONS SRS) for archiving in the EEF data archive (managed by FFT Education) and will include data only individually identifiable to the Department for Education.

Both the evaluation and delivery teams will seek ethical approval for the study. The study will not commence until ethical approval is received.

How will the data be shared?

During the trial, data will be shared securely using a specialised software (Egress). Majority of the data that will be used for the purposes of the intervention and the evaluation will be stored and processed within Egress, with licence applications processed by LexiaUK under the terms of their [Privacy Policy](#).

Pupil names and date of birth will be shared with Elklan to enable them to collect outcome data. For research purposes, following the completion of the trial, the data will be shared with the EEF archive who will act as the data controller after that point. You can read the EEF's Data Protection Statement in the EEF's website.¹³

How will the data be kept secure?

The delivery and the evaluation teams have put appropriate security measures in place to collect and to keep personal data secure and to prevent any unauthorised access to or use of it. All data will be collected and stored in accordance with the Data Protection Act (2018) and GDPR (General Data Protection Regulation) requirements. Whilst, in usual practice, some data from the Lexia software is stored in servers located in the United States, no data used for the evaluation will be saved on servers outside the EEA or the United Kingdom, pursuant to EEA-approved terms. Evaluation data will be stored on secure servers. Data transferred between the delivery and evaluation teams will be encrypted or use secure file transfer protocols, primarily through Egress. Please see LexiaUK's Privacy Policy for how software data is collected and processed.

What is the lawful basis we rely on in order to use and collect personal data and special category data?

The data collected for the Lexia project is used for the purposes of research. The lawful basis for LexiaUK and RAND Europe's use of that data under the GDPR is 'legitimate interest' as permitted under GDPR Article 6(1)(f) as the research will produce beneficial information for schools that will benefit children.¹⁴ The Delivery and Evaluation teams will use data collected through the NPD linkage for the purposes of research. The Evaluation team will use the basic identifiers (name and date of birth) to associate information with individuals to create datasets and that use will not affect the rights and freedoms of the children because this information will not be publicly connected to any other information. The test data will not be publicly available in association with the basic identifiers and therefore no risk of detriment exists to the children. Special category data is processed for the purpose of scientific research and archiving as permitted under GDPR Article 9(2) (j).

How will existing data be archived?

The Evaluation team will archive data pertaining to the Lexia trial as per the Table above into the EEF's archive in the Office for National Statistics Secure Research Service. All this data may potentially be linked to the NPD or to other data sources in the future for the purposes of research. When archived, this data has the potential to be used in further educational studies. No personal identifiers are submitted to the archive except for the Pupil Matching Reference, which is needed to link trial pupils with administrative data. The EEF archive has recently been transferred to the Office for National Statistics Secure Research Service, which significantly increases the security of the data. It can only be accessed by accredited researchers within the secure service. This fully complies with GDPR and the Data Protection Act 2018.

How long will we keep data?

Research data will be kept securely by the delivery team (LexiaUK and QUB), the evaluation team (RAND Europe), and the testing sub-contractor (Elklan) for the duration of the study. This data, including test outcome data collected by Elklan, will be deleted one year after the report has been published. The delivery team will keep anonymised data (i.e., with no identification of pupils or schools and no date of birth, just age in months) that cannot be linked to any individual for three years after the publication of research papers, which is the standard period required for academic publications. Additionally, pupil data and data on the delivery of the intervention will be deposited into the EEF's data archive in the Office for National Statistics Secure Research Service at the end of all of the research.

¹³ For more information about the EEF archive and Data Protection Statement, please see: https://educationendowmentfoundation.org.uk/public/files/Evaluation/Data_protection/Data_protection_statement_EEF_evaluations.pdf

¹⁴ For more information about legitimate interest, please see: <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/legitimate-interests/what-is-the-legitimate-interests-basis/>

What choices do parents have in our use of their child's data?

If parents do NOT want their child's data to be used, we will provide a withdrawal form for parents to return to their child's class teacher. During or after the intervention, participants may contact RAND Europe and/or Queen's University Belfast to request deletion of their personal data.

Should parents wish for their child's data to be withdrawn over the course of the evaluation, they may contact the RAND Europe evaluation team at lexia@randeurope.org, who will then instruct LexiaUK to proceed with data removal.

What are your rights?

The Evaluation and Delivery teams will operate in accordance with EU law including GDPR. You are provided with certain rights that you may have the right to exercise through us. In summary those rights are:

- To access, correct or erase your personal data
- To object to the processing of your personal data

Were you to request information from us, we will need to confirm your identity to ensure the security of your data. We will endeavour to respond within 30 days, but our response time may vary depending on the complexity of your request.

If you wish to exercise any of these rights please contact the RAND Europe Data Protection Officer by emailing REdpo@randeurope.org or in writing to Data Protection Officer, RAND Europe, Eastbrook House, Shaftesbury Road, Cambridge, CB2 8BF, UK. Once the project (Independent Evaluation of the Lexia® Core5® Reading Programme, 23035) is completed and data has been shared with the EEF archive, participants should contact the EEF at info@eefoundation.org.uk.

You also have the right to contact the Information Commissioner's Office in the UK if you have any concerns about the processing of your data by RAND Europe. You can visit ico.org.uk or email casework@ico.org.uk. Alternatively, you can write or telephone to Information Commissioner's Office, Wycliffe House, Water Lane, Wilmslow, Cheshire SK9 5AF.

Who can I contact about this project?

For questions about Lexia delivery, please contact Joanna O'Keefe of the delivery team at lexia@qub.ac.uk.

For questions about testing and evaluation, please contact Elena Rosa Speciani or Miguel Subosa of the evaluation team at lexia@randeurope.org.

WITHDRAWAL FORM

Independent Evaluation of the Lexia® Core5® Reading Programme

Please complete only if you do NOT want your child's data to be used in this research.

If you are happy for your child's data to contribute to this research and to be archived by the Education Endowment Foundation, then you do not have to do anything.

If you **do NOT** want your child's data to be used for this study or their data to be archived, please fill out the form below and return to your child's school.

☐ I, the undersigned, hereby do NOT give permission for my child's data to be used for the purposes of the Lexia® Core5® Reading Programme trial.

☐ I, the undersigned, hereby do NOT give permission for my child's data to be archived by the Education Endowment Foundation.

Child's full name: _____

School name: _____

Parent/carer's name: _____

Parent/carer's signature: _____

Date: _____

Appendix I: Randomisation code

Written in Stata 18

*load the data

use [file path]

*check how many schools have EIA status

tab eia

*62 schools with eia and 51 without so we should be looking for:

*31 schools with eia in treatment and 31 with eia in control

*25/26 without eia in treatment and 25/26 without eia in control

*we use the randtreat package to run the randomisation

ssc install randtreat, replace

*we use randtreat to generate a new variable "treatment" which is stratified by eia status (equal (non-)eia in treatment and control), creating two different groups (multiple(2)) and using a seed of 999 to ensure that this code generates the same assignment each time

*since we have an unequal number of schools taking part in the trial, we need to assign a misfit strategy to deal with the one school left out of the equal group assignment - here we assign the strata method which effectively randomly assigns it to one group in this context

randtreat, generate(treatment) setseed(999) strata(eia) multiple(2) misfits(strata)

*now we check to make sure we have equal assignment of eia and non-eia in both treatment and control

*first for schools in an eia

tab treatment if eia == 1

*31 in treatment 0 (control) and 31 in treatment 1 (treatment)

tab treatment if eia == 0

*25 in treatment 0 (control) and 26 in treatment 1 (treatment) - unbalanced because of 1 misfit

*pull in the school names for the dataset

merge 1:1 schoolid using [file path]

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>



**Education
Endowment
Foundation**

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

<https://educationendowmentfoundation.org.uk>

 @EducEndowFoundn

 [Facebook.com/EducEndowFoundn](https://www.facebook.com/EducEndowFoundn)