Journal of Applied Developmental Psychology 95 (2024) 101726

Contents lists available at ScienceDirect

Journal of Applied Developmental Psychology

journal homepage: www.elsevier.com/locate/jappdp

The benefits of computer-assisted learning for struggling readers in elementary schools in England

Maria Cockerill [a,*], Louise Tracey [b], Louise Elliott [c], Caroline Fairhurst [c], Laura Mandefield [c], Imogen Fountain [c], Sarah Ellison [d], Allen Thurston [a], Joanne O'Keeffe [a]

[a] School of Social Sciences, Education and Social Work, Queen's University Belfast, United Kingdom
[b] School of Education, University of Leeds, United Kingdom
[c] York Trials Unit, University of York, United Kingdom
[d] Department of Education, University of York, United Kingdom

ARTICLE INFO

Keywords:
Computer-assisted learning
Reading
Literacy
Randomized controlled trial
Implementation
Socio-economic disadvantage

ABSTRACT

Results of a randomized controlled trial in English elementary schools of *Lexia Core5 Reading* are presented here. The research assessed whether a computer-assisted learning program designed to improve reading outcomes for all readers, when delivered as a targeted intervention to struggling readers improves reading outcomes for these First Grade students. The program was delivered successfully across one school year. Analysis was undertaken for 620 students, from 57 participating schools with a mean average of socio-economic disadvantage above the national average. Positive *Effect Sizes* were observed of +0.08 overall and of +0.18 for low socio-economic status (SES) students. A larger study is warranted to ascertain generalizability to a larger population including in other grades.

Introduction

Reading proficiency is a challenge for children worldwide. The World Bank reported 70 % of 10-year-olds from low and middle-income countries struggle to read and understand a short, age-appropriate text (World Bank, 2022). Despite being a high-income country, 25 % of children in England failed to reach expected standards in reading at age 5–6, assessed through a phonics check (Department for Education, 2022) and 27 % failed to meet expected standards at age 11 (DfE, 2023). Furthermore, 41 % of all students did not meet expected standards aged 11 in reading, writing and maths (This is a combined measure of all three subjects, all of which require reading skills to understand the curriculum) (Department for Education, 2023a, 2023b). It is clearly important to support literacy in the early elementary stages to ensure children can read well as they age to prevent them from struggling academically throughout their school education and beyond. Evidence from the U.S. supports the need to intervene early, as children who do not meet expected standards entering Third Grade also continue to struggle to achieve appropriately throughout the school years (Feister, 2013). This is evidenced by National Assessment of Educational Progress (NAEP) figures which indicate that 37 % of fourth-graders performed below NAEP Basic in reading, an increase from 2019 where the figure was 33 % (NAEP, 2022). In England, we also know that poor reading attainment is a particular issue facing low SES children. Children here also struggle with reading proficiency, and this is most acute for children from poorer backgrounds, as 49 % of low SES students fail to reach expected reading standards aged 11 years (Department for Education, 2023a, 2023b).

To support children's reading, policy makers in England recommend classroom level teaching which integrates phonics, fluency and comprehension skills, as well as additional targeted support for children struggling with their reading proficiency (Education Endowment Foundation, 2017). Using technology-based learning with struggling readers, particularly for those from high poverty backgrounds has been found to be effective in improving children's reading (Higgins, Xiao, & Katsipataki, 2012). Using technology including computer-assisted learning (CAL) educational technology programs is popular with schools in England as they include personalised and often independent learning matching student reading abilities and the capacity to adapt and focus on a child's specific needs. Systematic review evidence from the US finds educational technology applications including computer-assisted learning an effective means of improving reading outcomes

* Corresponding author.
E-mail address: maria.cockerill@qub.ac.uk (M. Cockerill).

https://doi.org/10.1016/j.appdev.2024.101726
Received 5 March 2024; Received in revised form 11 November 2024; Accepted 17 November 2024
Available online 24 November 2024

for struggling readers (Cheung & Slavin, 2012). With regards to the Lexia's CAL programs, one of which is being studied here, Slavin, Lake, Chambers, Cheung, and Davis (2009) report high levels of effectiveness for younger readers and a recent RCT undertaken in US schools of a Lexia computer-assisted program, *Lexia PowerUp Literacy*, supports its use with students aged 13–14 (Hurwitz & Macaruso, 2021). Small scale studies in England also indicate these computer-assisted programs, in this case *Lexia Core5 Reading*, can help children's reading although the implementation of the program was only undertaken for 8 weeks (O'Callaghan, McIvor, McVeigh, & Rushe, 2016).

The present study reports the results of an efficacy randomized controlled trial (RCT) of the current *Core5* program in elementary schools in England, to extend the existing evidence and assess the extent to which previous findings are replicated at greater scale in this context with struggling readers in First grade. This study contributes to the literature as it is the first study of *Core5* in the UK to use randomization at the student level with a large sample size across multiple sites to test whether it is implementable at this scale, and to test its effectiveness in improving reading outcomes for children in elementary schools. Given the current reading attainment struggles faced by children from socio-economic disadvantage in particular, this study also tests the potential of this technology assisted programme to improve reading outcomes for this group of children in England.

Background

Activity theory: A theoretical lens for technology-based learning

Activity Theory provides a theoretical lens through which to view human–computer interaction. This theory evolved from the ideas of Vygotsky in the 1920s (Engeström, 1990, 2000; Vygotsky, 1978), where a subject (an individual) acting within a community uses tools to achieve a purpose or object. The object is the motivation for the activity, and the activity is mediated by a tool or artefact (for example, technology). The process of the subject (eg. a student) working towards an object (eg. Skilled reading) using a tool (technology) brings about an outcome (eg improved attainment in a particular subject) (McAvinia, 2016). Engeström (2000) helpfully articulates the 'activity system' whereby the subject, working to achieve the object by using a tool, acts within a community which includes rules and divisions of labour for their functioning.

In the 1980–90s researchers from Scandinavia and the U.S. suggested that by framing human-technology interaction within a larger context of purposeful human activities, the theory makes it possible to reach a deeper understanding of technology's benefit for people (Kaptelinin & Nardi, 2018). Kaptelinin and Nardi (2012, 2018) suggest that Activity Theory (AT) still offers researchers useful insights and guidance in studies of human-technology interaction, particularly where the focus is on the use of technology in relation to important human concerns. For example: Woll and Bratteteig (2018) used Activity Theory to understand technology-mediated elderly care and how it might enhance independence; Cornet, Voida, and Holden (2018) used AT to understand the benefit of computer-supported healthcare systems which help those with chronic health conditions; and Ferreira Lemos, da Cunha, and Júnior (2018) used Activity Theory to understand how human-technology interaction using social media was capable of mobilising school students in social movements in Brazil. The Activity Theory lens has also been used to view innovations in education. For example: Issroff and Scanlon (2001, 2002) used Activity Theory to explore technology use in higher education and suggest its potential to highlight problematic features of the learning and teaching setting; Holen, Hung, and Gourneau (2017) and also Al-Huneini, Walker, and Badger (2020) looked at the introduction of individual laptops in rural schools, albeit in different countries, and identified its benefits and challenges; and, Zheng, Kim, Lai, and Hwang (2019) explored flipped learning and made recommendations for more effective implementation of this learning.

In the case of CAL in schools, the student would be the subject learning within a community (including rules for teacher support with divisions of labor between student and teacher) using the interaction medium of a computer program, with the desired outcomes of improving their reading skills. Activity Theory is not being used in this study as a systematic framework for analysis, but instead as a lens through which to view the mechanisms of change which gave rise to the findings, and to discuss the potential of using CAL to improve children's reading and to highlight problematic features.

Using educational technology to improve outcomes for children

Evidence suggests that technology used in education can provide a medium through which students can practice skills and consolidate their knowledge, either through independent work or when supported by a teacher (Lewin, Smith, Morris, & Craig, 2019). Researchers point to an association between educational technology use and educational attainment (Wainer et al., 2008) with positive findings for smaller and more intensive interventions (Liao & Hao, 2008). A synthesis of 48 studies about the impact of the use of digital technology in schools suggests digital equipment, tools and resources can at times improve attainment outcomes (Higgins et al., 2012). However, although the typical overall effect size in the general analyses of the impact of educational technology on learning is between +0.3 and + 0.4, the range of effects is very wide (−0.03 to +1.05). Higgins et al. (2012) recommends examining the differences between educational technologies and how they are used to best effect and considering educational technology as a catalyst for change to make teaching and learning practices more effective.

Although educational technology is found to support student learning, educational technology itself does not guarantee effective learning. Supporting strategies are found to be important to the learning process using educational technology (Chen, Wang, Kirschner, & Tsai, 2018). For example, Vogel, Wecker, Kollar, and Fischer (2017) find positive effects of educational technology (ES + 0.95) when supported by teacher facilitation, instruction and guidance. Instruction and guidance, including teacher facilitation to scaffold learning, can reduce students' off task behaviours. Teacher facilitation takes place when the teacher becomes a facilitator as well as a knowledge expert, to support and guide the learning process (Hsieh & Tsai, 2012; Michinov & Primois, 2005). As teacher facilitation includes cognitive and affective strategies, and helps students articulate ideas and engage in meaningful interactions, the cognitive strategy promotes student's knowledge construction helping them focus on the learning task or topic, and the affective strategy improves student's motivation by the provision of positive feedback. It is perhaps for this reason that Chambers et al. (2008) find the combination of computer-based activities combined with direct teacher instructional support, helpful for improving student's learning.

Reading development and the contribution of a computer-assisted learning program

It is widely accepted that five areas of reading should be addressed during reading instruction: phonemic awareness (knowledge about the units of sound); phonics (knowledge of letter sound correspondence); fluency (the ability to accurately and quickly decode text to facilitate comprehension); vocabulary (the ability to understand and articulate the meaning of words); and comprehension (the ability to make literal and inferential meaning from what is read) (National Institute of Child Health and Human Development, 2000). For example, systematic phonics teaching is associated with better progress in reading accuracy, evidenced at all levels of ability (Torgerson, Brooks, & Hall, 2006). Evidence syntheses recommend approaches with phonology emphasis, with highest effects found for combinations of real reading with systematic phonics (Slavin et al., 2009).

A computer-assisted learning program such as *Core5* used in the present study includes various activities that teach phonetic word-attack strategies for automaticity in word recognition, yet also include the other elements necessary for reading instruction.

It is suggested that educational technology in the form of CAL programs may be part of the long-term solution for student phonics and more generally reading support, given its capacity to provide specialised instruction and practice for relatively low cost with the potential for relatively high, consistent fidelity (Torgesen, Wagner, & Rashotte, 2010). Researchers corroborate this view, that CAL programs focused on improving reading development can improve educational attainment. Results from a large-scale trial in England evaluated by McNally, Ruiz-Valenzuela, and Rolfe (2018) reports a particular benefit for children identified in the category of low SES (using the proxy of eligibility for Free School Meals – FSM). McNally et al. (2018) found positive reading outcomes of a higher magnitude for this group of students (ES + 0.368) compared to the positive outcomes found for all children (ES + 0.138) using CAL to improve their literacy. These authors also tested the same technology-delivered program against a control group using traditional delivery through teacher instruction. McNally et al. (2018) found no additional benefit of literacy instruction delivered to students via technology when compared to traditional teacher-led delivery, which suggests the content more than the delivery mechanism is what leads to improvement. Nevertheless, Lewin et al. (2019) remind us that technology does offer an alternative and effective means of delivering literacy content to children, enabling them to practice their language skills with little adult support. In the context of schools who must support students who struggle with literacy in classrooms where teachers cannot always offer the individualized teaching time students require, additional technology delivered interventions can be a pragmatic and effective solution to improving reading outcomes.

The evidence base of Lexia Core5 Reading and the importance of the current study

The Lexia program was originally developed in the USA in 1984 for children with dyslexia. Benefits of using a blended approach which includes CAL and teacher resources to support usual practice which the Lexia program provided are reported from a quasi-experimental study with 106 struggling readers aged 6–7 in the UK (McMurray, 2012). The current version of the Lexia program, *Core5,* was designed for wider use and is now available through LexiaUK in England and recommended as an additional intervention-style component of any literacy curriculum, particularly for struggling readers. The program includes a blended approach of computer-based activities for students, combined with the provision of paper-based resources to supplement learning delivered by school staff where necessary. The *Core5* program has received a 'strong' rating in the US as outlined by the federal law under the Every Student Succeeds Act, based on evidence from a randomized controlled trial undertaken by Hurwitz and Vanacore (2022) which included 116 students in five schools with positive findings with an Effect Size of +0.23 (Evidence for ESSA, 2024). In the UK, Evidence for ESSA (2024) also supports *Core5* with a 'moderate' rating based on positive impact for children between Kindergarten and Grade 1 in a study which detected an Effect Size of +0.16 using a quasi-experimental design across 167 schools in the US (Hobbs, 2016). Despite its size, the authors warn against the strength of these findings owing to a lack of random assignment to groups, suggesting the results should be seen as probable outcomes only (Hobbs, 2016). Other studies have also detected positive findings from the use of the *Core5* program: Wilkes et al. (2020) supports the value of using the *Core5* program in early elementary grades, although the findings from this study are also based on results from a quasi-experimental design with no random allocation. O'Callaghan et al. (2016) undertook a small randomized trial with 98 4–6 year-olds in the UK which boosted student's phonological skills, with positive effect sizes detected for blending (ES + 0.06) and non-word reading (ES + 0.07). As Evidence for ESSA (2024) suggests, the study design with a short implementation length of eight weeks was too short a time to yield robust findings.

Studies have also shown the potential benefits of using *Core5* with children from low SES backgrounds who struggle with reading. Macaruso, Wilkes, Franzén, and Schechter (2019) undertook a three-year longitudinal study of *Core5* with 68 children from low socioeconomic status (SES) backgrounds, following them from Kindergarten to Grade 2. Findings showed that children who engaged in the program made greater than expected progress, particularly in Kindergarten. However, the study findings were limited by a lack of a control group. Another study undertaken by Schechter, Macaruso, Kazakoff, and Brooke (2015) included a randomized design involving one school with two first and second grade classes with 83 students in total, where one class was allocated to treatment and one to control in each year group. The study found that instruction through the *Core5* program yielded an Effect Size of +0.53 in comprehension for treatment class versus the control class. Although these findings are drawn from a small sample of students with randomization at the class rather than individual student level, the study supports the potential of this digital program to improve reading outcomes for low SES children. The findings from these studies suggest that the *Core5* program should be tested at greater scale using individual randomization including a larger sample of students, to yield more robust results.

The present study contributes to this body of evidence by testing the computer-assisted program *Core5,* delivered by LexiaUK*,* using a two-armed efficacy randomized controlled trial (RCT). Given that allocation was undertaken at the individual student level with a large sample size of 697 students of whom 265 were low SES, across multiple sites ($n = 57$ schools) the study was well placed to test whether the program is implementable at this scale, and to test its effectiveness in improving reading outcomes for children in elementary schools (including a particular focus on low SES children). Student allocation was undertaken within schools with students allocated to receive the *Core5* program or not. This study provides a unique opportunity to evaluate the program using a larger scale RCT within the UK context than previously attempted, using standardized measures that assess all round reading ability.

Lexia Core5 Reading: The intervention used in the present RCT

In line with recommended best practice (Department for Education, 2013; Department for Education, 2023a, 2023b; National Institute of Child Health and Human Development, 2000), *Core5* supports classroom teaching as it is designed to supplement usual teaching, providing children who require it with individualized reading practice in six areas: phonological awareness, phonics, structural analysis, automaticity, fluency, vocabulary, and comprehension. The computer tracks student progress as they learn using the computer program independently and provides extra practice on areas of difficulty. Staff give initial guidance on program use, and teach or reinforce elements where necessary, and primarily oversee and monitor student independent online use. In this way, *Core5* is an important resource for schools as they support students who struggle with literacy in classrooms where teachers find it difficult to offer the individualized teaching time required by struggling students without the aid of the online program.

Staff training includes three (1 h) online webinar-style-training sessions per school for a senior leader and day-to-day users (teacher/teaching assistant):

- Initial Training: Lexia Overview/Getting Started.
- Reports Training: review assessment data/implementation/progress.
- Review progress, data and discuss next steps.

Ongoing technical support and support for teachers, including online web-based training is available for schools.

The intervention was used in the present study with selected students in Grade 1 (aged 6–7 years old), who were identified as struggling readers. Within each school half the selected students were randomly allocated to receive the intervention in addition to the core national curriculum (Department for Education, 2013) or to act as control and continue with usual teaching and learning. Schools received the program software including off-line paper-based resources for use by teachers when necessary, teacher dashboard reports and student reports to inform schools of student progress and usage were available for schools to download. Students used the program on a tablet or PC outside of the classroom as is usual practice for intervention delivery in elementary schools in England. The intervention was scheduled for use four times per week for 30 min (including 10 min set-up time) over the academic year. For compliance, students must have completed a minimum of 60 min (excluding set-up time) per week for at least 12 non-consecutive weeks. Given that students require a designated login to use the online program, there was no risk of contamination of the control group as these students were unable to access the program during the trial. During the trial the control group continued with usual teaching and learning, and the treatment group received usual teaching in addition to the *Core5* intervention. Usual, or 'universal', teaching for Grade 1 students in England is in line with the national curriculum guidance for Reading which includes practice in the following areas: phonological awareness, phonics, structural analysis, automaticity, fluency, vocabulary, and comprehension (Department for Education, 2013). For example, the national curriculum requires children be taught to: apply phonic knowledge and skills as the route to decode words; respond speedily with the correct sound to graphemes for all 40+ phonemes; read accurately by blending sounds in unfamiliar words; read common exception words; read words of more than one syllable; read words with contractions; read aloud accurately and re-read these books to build up fluency; make inferences, predictions and summarise what they have read (Department for Education, 2013). In this way, struggling readers in the treatment group received a blended learning approach which included usual teaching and the *Core5* program.

Lexia Core5 Reading: Theory of Change (ToC)

Fig. 1 illustrates the ToC. By providing a structured computer assisted intervention for struggling readers in lower elementary years, and training for lead and delivery staff for its successful implementation including in the use of resources for bespoke student support, it is proposed that the processes underpinning the *Core5* program are capable of leading to improved reading outcomes for students. This assumes that the training for schools impacts on the professional action of delivery staff to facilitate the implementation of the program, as they are trained to use the teacher dashboard, to monitor student progress by downloading student reports, and to intervene through teacher supported instruction using downloadable physical materials to consolidate student learning where necessary at any given point in the program. As a result, it is projected that student participation in the *Core5* program which includes a blended model of CAL and where necessary teacher instruction leads to improved reading.

Research questions

The study addresses the following research questions outlined below, with the purpose of contributing to the literature about CAL technology use, and more specifically to contribute to the evidence base about the effectiveness of the educational technology program *Core5*. The study aims to build on the existing evidence about the benefits of using *Core5* through a larger scale RCT design than has been previously attempted, focusing on its use as an intervention in addition to usual core curriculum learning with children in England who struggle with reading,

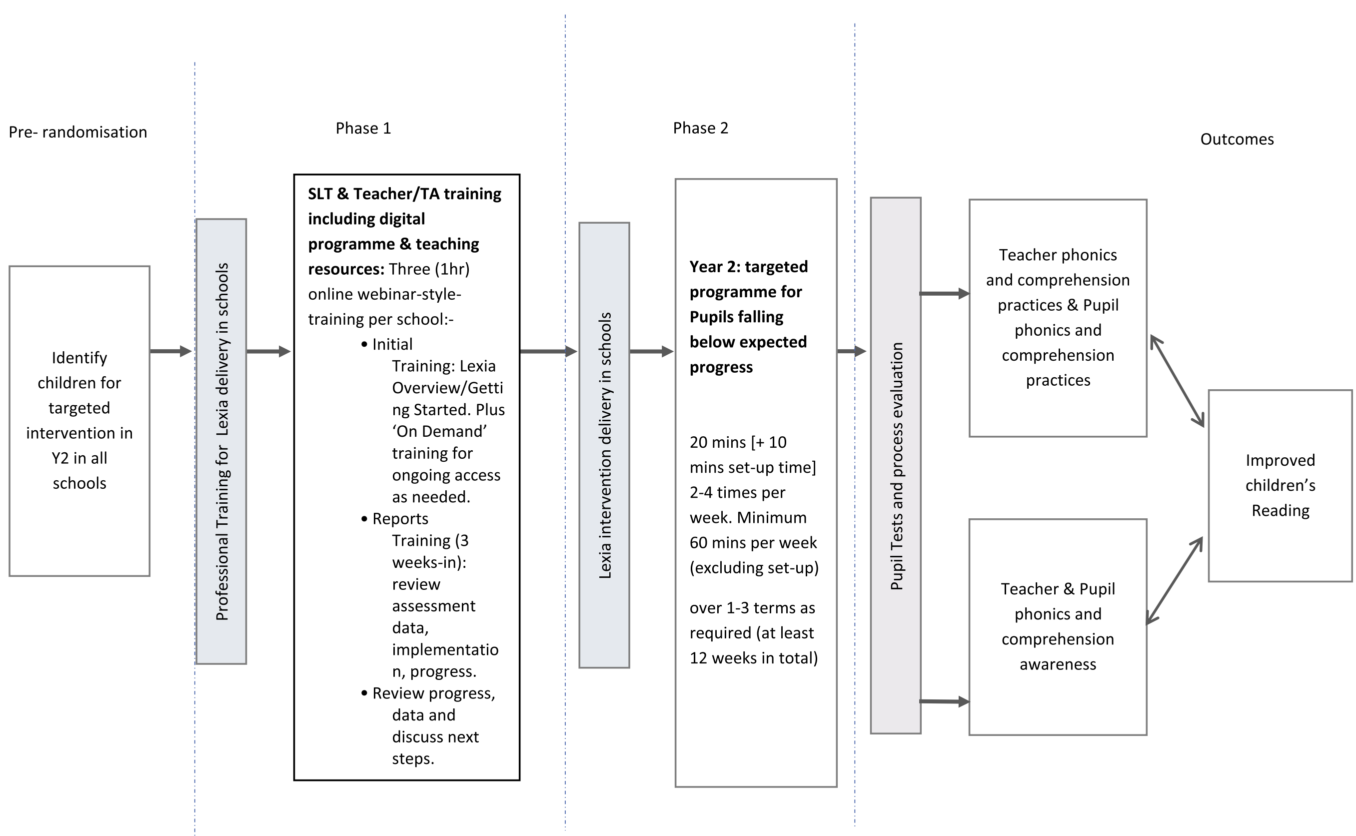

Fig. 1. Logic Model.

including with low SES children. Activity Theory provides a helpful 'activity system' frame through which to consider fidelity of implementation, to explore the mechanisms of change, including delivery structures and teacher facilitation and support.

a) Can *Core5* be implemented with fidelity in elementary schools in England at greater scale than previously tested?
b) Can *Core5* improve reading skills for struggling readers during First Grade when compared to a control group?
c) c) Can *Core5* improve reading skills for low SES students during First Grade when compared to a control group?

Method

The study consisted of a RCT of the computer-assisted program *Core5* in elementary schools in England across one academic year (2018–19) (Tracey et al., 2022). The trial was commissioned and funded by the Education Endowment Foundation (EEF), and the protocol is published on their trial website (EEF, 2021). Ethical approval for the study was granted by the Research Ethics Committees of the School of Social Sciences, Education and Social Work at Queen's University Belfast, Northern Ireland, and the Department of Education Ethics Committee at the University of York, England. In addition, participating school headteachers approved the intervention of the trial and data collection through a signed memorandum of understanding (MOU). Parents or guardians of selected students in each school were also provided with information in the form of a letter that informed them that the *Core5* trial was taking place in their child's school, and provided them with information about the project, including data collection for participants, the data privacy policy and the option to withdraw their child from data collection activities.

Sample

The trial was designed to recruit 57 schools given the capacity of the delivery team. It was felt important that eligibility at the child-level was determined independent of teacher judgement to ensure consistency across participating schools. Consequently, children were independently assessed at the end of the academic year prior to entering Grade 1 using the Word Identification, Word Attack and Passage Comprehension subtests of the Woodcock Reading Mastery Tests, Revised Normative Update (WRMT-R/NU) to determine inclusion in the trial (as well as providing a baseline measure). It was originally envisaged that students with a standard age score (SAS) of less than or equal to 85 (classed as below average: Castro & Robnolt, 2016) in any of the three sub-tests of the WRMT-R/NU would be eligible to participate in the study. Assuming an average of 56 students in a year group in the school (28 per class, two-form entry) and based on a similar study using a similar assessment (Tracey, Chambers, Slavin, Hanley, & Cheung, 2014), it was estimated that there would be an average of nine struggling readers per school. Assuming a correlation of 0.6 between the baseline and post-test of the WRMT and 80 % power, the minimum detectable effect size (MDES) would be 0.20 adjusting for 10 % attrition at the student level (StataCorp, 2017).

Recruitment and inclusion

Recruitment was conducted in the North East, Yorkshire and Humber, the South West, the North West and the Greater London area regions of England. Recruitment focused on schools with an above national average of students who were eligible for free school meals (FSM), (a proxy for socio-economic disadvantage). To be eligible to participate schools needed to:

- have approximately 50 students per year group (i.e. be at least two-class entry)
- not be involved in any other EEF trial focusing on Grades 1–3
- not currently using *Core5* (including in the previous 12 months)
- have the necessary IT to support implementation
- be willing to implement the intervention with respect to the random allocation (i.e. to only deliver the intervention to those students assigned to the intervention condition).

Schools then selected children due to enter Grade 1 classes the following autumn who were in the lowest half of the year group as determined by their prior attainment in reading ability. Schools were eligible to be randomized (and thereby included in the trial) after these students had been assessed at baseline (during the end of the summer term before they entered Grade 1), schools had signed the MOU and Grade 1 teachers had completed an on-line pre-randomization survey.

Participants

All schools who had met the criteria above were included in the trial. Children were selected for inclusion based on their results from the baseline assessments. Schools indicated that they had the capacity to deliver the intervention to six or seven students in Grade 1, which was higher than expected. Therefore, it was decided to include the lowest attaining (at baseline) 12 students in the randomization rather than using a cut-off value from the baseline assessment as originally intended. If there was more than one student with the same score as the 12th student, then up to 14 students were included in the randomization. The implication of using a relative cut-off was that some students would not generally be regarded as 'struggling readers' but fewer than expected children were assessed at reading below the established 'cut-off' level. Hence, this was balanced against having fewer students included in the trial.

Randomization

Students from all schools were randomized at the same time to ensure allocation was concealed from schools. Randomization was completed before the end of the academic year and before the commencement of intervention delivery at the start of the following academic year. To ensure a maximum of seven students were randomized to the Lexia intervention in any one school, block randomization with a fixed block size of 2 was implemented. Students were randomly allocated in a ratio of 1:1 to receive either the intervention or usual teaching. A statistician independent to the trial generated the allocation schedule, using STATA (StataCorp, 2017).

Measures

Baseline

Woodcock Reading Mastery Tests-Revised/Normative Update (WRMT-R/NU; Woodcock, 1998) Composite Measure (Word Identification, Word Attack, Passage Comprehension).

The WRMT-R/NU is a standardized measure suitable across a wide age range (5 years – 79 years 11 months). Overall, it reports an internal reliability coefficient 0.68 to 0.99 and a standard error measurement (SEM) of 2.0 to 6.7 (W scale units) for the G or H version of the test (Woodcock, 1998). When normed it correlated highly with the PIAT Reading (0.87 to 0.87).

Three sub-tests were administered: Word Identification, Word Attack and Passage Comprehension. These measure word recognition, word decoding and reading comprehension, respectively. Composite standard age scores were calculated where a student had a valid test score for at least two out of the three subtests and scores ranged from 74 to 166. The initial intention was to use the Woodcock Reading Mastery Tests (3rd ed) (WRMT-III; Woodcock, 2011) for baseline testing (see below for further details). However, due to issues with delivery of the WRMT-III, an older version of the test was administered. This was deemed suitable given that there is a high level of correlation between the two assessments. The correlation for the Basic Skills cluster between the two editions is 0.78 for K-G6 and for the Reading Comprehension cluster, 0.85 (see below for further details on clusters). As all students were assessed with the WRMT-R/NU at baseline the scores were considered comparable across the sample and therefore suitable for the baseline assessments. This composite score formed the baseline for the post-test WRMT-III composite score, the Oral Reading Fluency subtest and the Key Stage 1 reading scores at post-test.

Word identification. Word Identification measures word recognition, although there is no assumption that the child knows the meaning of the word. This assessment of the WRMT-R/NU scores from 70 to 159. The correlation with the Word Identification subtest of the WRMT-III is 0.75.

Word attack. Word Attack is designed to measure word decoding using 'nonsense' words to simulate a child encountering an unknown word in real life. It scores from 81 to 163. The correlation with the Word Attack sub-test of the WRMT-III is 0.73.

Passage comprehension. Passage Comprehension is designed to measure reading comprehension. During administration the child is requested to read a sentence or short passage and identify the missing word. Scores range from 70 to 175. The correlation with the Passage Comprehension sub-test of the WRMT-III is 0.79.

Outcomes

Woodcock Reading Mastery Tests (3rd Ed) (WRMT-III) (Woodcock, 2011) Composite Measure (Word Attack, Word Identification, Oral Reading Fluency, Passage Comprehension).

The WRMT-III is a standardized measure suitable across a wide age range (4 years 6 months to 79 years 11 months). The primary outcome measure used for Grade 1 students was a composite of the standard age scores of four subtests of the WRMT-III (Word Identification, Word Attack, Passage Comprehension, and Oral Reading Fluency), a paper-based measure with high reliability. The WRMT-III subtests are designed to measure word recognition (Word Identification), word decoding (Word Attack), reading comprehension (Passage Comprehension), and reading fluency (Oral Reading Fluency). There are no existing reliability and validity statistics for this composite measure. However, the Word Attack and Word Identification subtests together form the Basic Skills cluster of the WRMT-III which has a test-retest reliability for Grades K-2 of 0.90. Similarly, the four subtests all contribute to the Reading Comprehension Cluster of the WRMT-III which has a test-retest reliability for Grades K-2 of 0.85. The WRMT is highly correlated for pre-K-Grade 6 with the Kaufman Test of Education Achievement, Second Edition (KTEA-II; Kaufman & Kaufman, 2004) with correlations of 0.76 for both the Basic Skills and the Reading Comprehension Clusters,[1] and with the Wechsler Individual Achievement Test, Third Edition (WIAT-III; Wechsler, 2009) with correlations of 0.89 for Basic Skills and 0.78, for Reading Comprehension.[2] It should be noted, however, that the Reading Comprehension Cluster of the WRMT-III includes the four subtests used in this study and the Word Comprehension subtest, which was not used in this trial. The Word Comprehension subtest was not administered at baseline due to the amount of time testing on this scale would have taken (and additional participant burden) and because it would replicate some of the skills assessed in the Passage Comprehension subtest. The four subtests selected were judged to assess the key areas in which readers typically struggle and those that *Core5* specifically targets. Consequently, this composite measure was deemed to be appropriate for this trial. The composite score was constructed by summing the scores of the subtests, where a student had four valid results. Scores range from 231 to 580, with higher scores indicating higher proficiency in reading ability and reflects overall reading ability (the outcome measure). It should, however, be noted that Oral Reading Fluency was not available in the WRMT-R/NU so was only assessed at post-test. The reliability and validity of the individual subtests are provided below.

Word identification. See description above. The test-retest reliability for the Word Identification sub-test of the WRMT-III for Grades K-2 is 0.95. It is highly correlated (Pre-K to Grade 6) with the Letter & Word Recognition sub-test of the KTEA-II (0.75) and the Early Reading skills subtest of the WIAT-III (0.86). Scores range between 55 and 145 (see Table 1 for all subtest scoring ranges).

Word attack. See description above. Test-retest reliability of Word Attack in the WRMT-III for Grades K-2 is 0.89. It is correlated (Pre-K to Grade 6) with the Nonsense Word Decoding sub-test of the KTEA-II (0.74) and the Pseudoword Decoding subtest of the WIAT-III (0.79). Word Attack scores have a range of 64–145.

Passage comprehension. See description above. Test-retest reliability for grades K-2 is 0.80. It is correlated (Pre-K to Grade 6) with the Reading Comprehension sub-tests of the KTEA-II (0.72) and the WIAT-III (0.62). Passage Comprehension scores between 57 and 145.

Oral reading fluency. Oral Reading Fluency is designed to measure reading fluency. The child is required to read a short passage (80 words for Grade 1 students) during which their fluency is assessed. It has a test-retest reliability for Grades K-2 of 0.76. It is correlated (Pre-K to Grade 6) with the Decoding Fluency sub-test of the KTEA-II (0.67) and the Oral Reading Fluency subtest of the WIAT-III (0.85). Scores range between 55 and 145.

Key Stage 1 (KS1) reading raw scores. In England, prior to 2024, there was an annual requirement for all students to sit nationally administered curriculum assessments (SaTs) in Reading, English and maths at the end of KS1 (Grade 1). These assessments are designed to measure student performance, and allow for comparison within and across schools and over time (Standards & Testing Agency, 2020). The results are standardized although for the purposes of this evaluation we used the KS1 raw reading scores. They are scored between 0 and 40.

In all outcomes, a higher score indicates a better outcome.

Data collection

At the end of the academic year before students entered Grade 1, identified participants were independently assessed by the administration of the WRMT-R/NU by assessors recruited and trained by the evaluation research team. Administration of the WRMT-III at post-test (post-intervention, 12 months after the pre-test WRMT-R/NU) was again conducted by trained administrators who were blind to group allocation to avoid potential for ascertainment bias. The post-test was administered post-intervention, 12 months after the pre-test WRMT-R/NU, in the last months of the academic year when children were in Grade 1. The KS1 reading tests are national curriculum tests taken during the summer term at the end of Grade 1. Participating children's KS1 raw reading scores were provided to the research team directly by schools.

Analysis plan

The analysis was pre-specified in a statistical analysis plan (Fairhurst & Rex, 2019) and was performed using R statistical software (v3.5.3) (R Development Core Team, 2011). Analyses followed the intention to treat

[1] Correlation with KTEA-II Composite Reading (Kaufman & Kaufman, 2004).
[2] Correlation with WIAT-II Composite Basic Reading and Composite Total Reading, respectively (Wechsler, 2009).

Table 1
Primary and secondary outcome measures.

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

(ITT) principle unless stated otherwise and estimates of the effect are presented with corresponding 95 % confidence intervals and *p*-values.

A summary of school characteristics has been presented. Student characteristics and measures of prior attainment by trial arm both as randomized and as included in the primary analysis are summarised. A Hedges' g effect size and 95 % CI are detailed for the unadjusted difference between arms on the pre-test scores. Intra-cluster correlations (ICCs) and pre- and post-test score correlations are reported.

Primary analysis

The primary outcome (WRMT-III composite reading score) was analysed using a linear mixed effects model at the student level with group allocation and measure of prior attainment included as fixed effects and school as a random effect.

Secondary analysis

The secondary outcomes were analysed using the same model as the primary outcome using an appropriate pre-test WRMT-R/NU score as the measure of prior attainment.

Analysis in the presence of non-compliance

The program was delivered for 24 weeks (although schools were permitted to keep using the online components with the intervention group beyond this initial delivery period providing the licences were used for these students only). Schools were expected to arrange four sessions lasting 30-min each week (including ten minutes for set-up).

Via the MOU and a data sharing agreement, participating schools provided opt-in consent for the evaluation team to access the data files produced by, and for, schools via the software (only one school did not agree to this aspect of the evaluation). The *Core5* program software provided reports relating to implementation, fidelity and dosage at the school, class and individual level, including frequency and length of time the child was logged into the Lexia ILS program, patterns of usage and the areas of the adaptive program that demonstrated the greatest amounts of progress over time. Data on *Core5* use for each individual student, including date and duration of the session, was exported from the computer program.

To examine the effect of compliance to the intervention a Complier Average Causal Effect (CACE) analysis for the primary outcomes was conducted. Three separate measures of student-level compliance were defined:

- a binary variable (complied or not). A complier was defined as those who completed at least one-hour (60 min) a week (excluding set-up time) for at least 12 (not necessarily consecutive) weeks;
- a binary variable (complied or not). A complier was defined as those who completed at least 12 h (720 min) in total; and
- a continuous variable, defined as the total number of hours of the intervention received.

A two-stage instrumental variable (IV) analysis was used with random group allocation as the IV (Dunn, Maracy, & Tomenson, 2005).

Missing data analysis

A summary of missing baseline and outcome data has been presented by trial arm and, if available, reasons for missing data provided. To identify variables that were associated with missing primary outcome data a mixed effect logistic regression model was used including group allocation and WRMT-R/NU pre-test scores as covariates. Multiple imputation by chained equations (MICE) was also used to evaluate the impact of missing data on the results of the primary analysis. Pre-test WRMT-R/NU scores, age, gender and FSM status were included as predictor variables, with a 'burn-in' of 150 and 30 multiply-imputed data sets (Azur, Stuart, Frangakis, & Leaf, 2011).

An additional sensitivity analysis was carried out using multilevel imputation which required the 'jomo' package in R (Quartagno & Carpenter, 2016). The result was compared to the MICE analysis for any major differences.

Subgroup analyses

An interaction term between FSM status and random group allocation was included in the primary analysis model to assess the effect of the intervention on students who were eligible for FSM. The primary analysis was also restricted to the subgroup of students who were ever eligible for FSM.

Protocol

A process evaluation was conducted alongside the impact evaluation. This paper does not report on the process evaluation, except in respect of naturally occurring data about usage to establish fidelity of implementation relating to dosage (see above), alongside additional data relating to implementation collected via a survey administered on-line to participating Grade 1 teachers. As indicated above, compliance was defined as completing at least one hour a week for a minimum of 12 (not necessarily consecutive) weeks or completing a minimum of 12 h over a 24-week period.

Results

Participant flow through the study

Fig. 2 depicts the flow of schools and students through the trial. Of the 437 schools approached to take part in the study, 162 (37.1 %) responded to the initial contact and were sent an information sheet and Expression of Interest (EOI) form. Seventy-eight schools returned an EOI and these were subsequently screened to determine eligibility. Five schools were excluded as they were currently using, or had recently used, the Lexia program. The 73 eligible schools were sent the evaluation MOU, 59 of which returned it, although two schools later withdrew prior to randomization.

A total of 697 students were randomized in 57 schools (348 intervention, 349 control). All 697 students completed a baseline WRMT-R/NU test and 620 (89.0 %) had complete primary outcome data (310 intervention, 310 control). For those who had incomplete primary outcome data; 20 of 697 (2.9 %) student assessments were incomplete, 43 (6.2 %) students had left the school prior to post-testing, six (0.9 %) were absent on the day of testing, two (0.3 %) withdrew from the study, and for six students (0.9 %) the reason for incomplete primary outcome data is unknown.

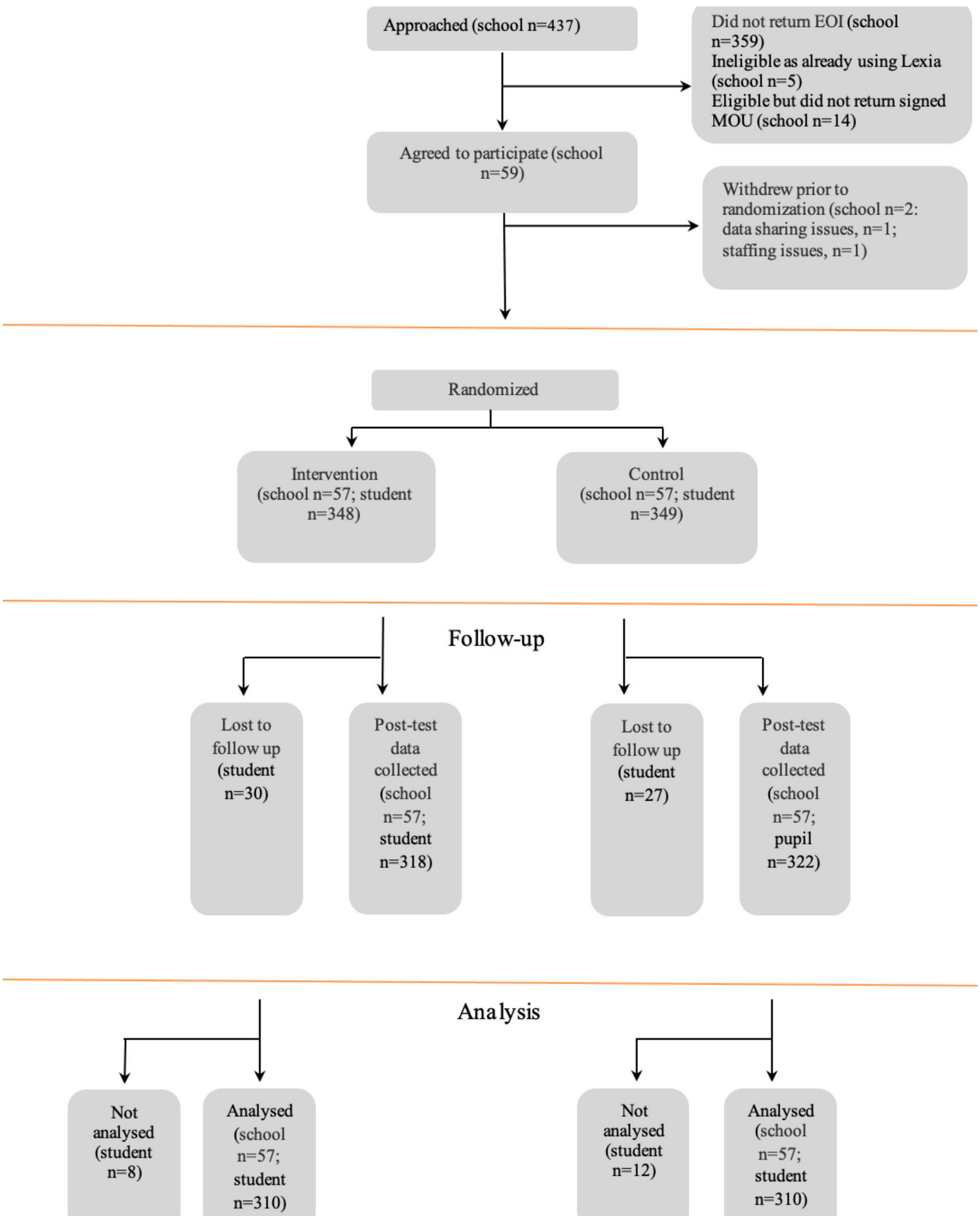

Fig. 2. School and participant flow chart.

Sample characteristics

School characteristics are presented in Table 2. As randomization took place within school at the student level, school characteristics are presented overall and not by group. Our sample is representative of the national level-means also shown in Table 2.

The baseline characteristics of the 697 randomized students and the 620 analysed students are presented in Table 3. Similar student characteristics were observed across the two trial arms. Pre-test scores were similar across arms, demonstrated by the small effect sizes associated with the difference.

Main analysis

Primary analysis

The adjusted mean difference in WRMT-III composite reading score

Table 2
Characteristics of 57 recruited schools.

School level (categorical)		Overall	
	National level mean	N (missing)	Count (%)
School setting*			
Rural town and fringe	N/A	57 (0)	6 (10.5)
Urban city and town	N/A	57 (0)	19 (33.3)
Urban major conurbation	N/A	57 (0)	24 (42.1)
Urban minor conurbation	N/A	57 (0)	8 (14.0)
Type of school			
Academy Converter	31 %	57 (0)	14 (24.6)
Academy Sponsor Led		57 (0)	6 (10.5)
Community school	38 %	57 (0)	26 (45.6)
Foundation school	3 %	57 (0)	6 (10.5)
Voluntary aided school	17 %	57 (0)	2 (3.5)
Voluntary controlled school	11 %	57 (0)	3 (5.3)
School Ofsted rating at time of MOU signed**			
Outstanding	18 %	48 (9)	9 (18.8)
Good	69 %	48 (9)	35 (72.9)
Requires Improvement	11 %	48 (9)	4 (8.3)
Inadequate	3 %	48(9)	0(0)
School level (continuous)***		N (missing)	Mean (SD)
Number on role	282	57 (0)	418.1 (139.3)
Percentage ever on FSM	N/A	57 (0)	29.9 (14.2)
Percentage of SEN students with a statement or EHC plan	1.6	57 (0)	1.1 (1.2)
Percentage of SEN students with SEN support	12.6	57 (0)	13.7 (6.0)

Ofsted, Office for Standards in Education, Children's Services and Skills; FSM, free schools meals, SEN, special educational needs; EHC plan, Education, Health and Care Plan.

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

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

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

between intervention and control groups was 3.63 (95 % CI -1.34 - 8.57). The estimated Hedges' g effect size for the primary analysis was of 0.08 in favour of the intervention (95 % CI -0.03 to 0.18); however this difference was not statistically significant (p = 0.15) (Table 4). The school level ICC for WRMT-III composite reading scores was 0.063 (95 % CI 0.010–0.121). The correlation between pre- and post-test scores was 0.75.

Secondary analysis

No statistically significant differences were observed in any of the secondary outcomes (the four subtests of the WRMT-III and the KS1 raw reading scores; Table 4) with effect sizes ranging from −0.03 to +0.1.

Sensitivity analysis: MICE

Following MICE, the adjusted mean difference in WRMT-III composite reading score was 3.09 (95 % CI -1.95 - 8.14, p = 0.23, effect size +0.06, 95 % CI -0.04 - 0.16; Table 5). Similar results were observed in the multilevel MICE model: an adjusted mean difference in WRMT-III composite reading score of 3.70 (95 % CI -1.28 - 8.68, p = 0.15, effect size +0.07, 95 % CI -0.03 - 0.17).

Complier average causal effect analysis

Of the total number of students randomized to the intervention, 338/348 (97.1 %) had data on their use of the intervention; one school (comprising of ten participating students) did not provide consent for

Table 3
Balance at baseline in intervention and control groups for randomized sample and analysed sample.

	Randomized (n = 697)		Analysed (n = 620)	
Student level (categorical)	Intervention group (n = 348)	Control group (n = 349)	Intervention group (n = 310)	Control group (n = 310)
	Count (%)	Count (%)	Count (%)	Count (%)
Gender				
Male	214 (62.2)	198 (57.1)	195 (62.9)	179 (57.7)
Female	130 (37.8)	149 (42.9)	115 (37.1)	131 (42.3)
FSM status*				
Not eligible for FSM	216 (62.8)	210 (60.5)	196 (63.2)	191 (61.6)
Eligible for FSM	128 (37.2)	137 (39.5)	114 (36.8)	119 (38.4)
Student level (continuous)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Age (years) at baseline assessment	6.36 (0.30)	6.39 (0.30)	6.36 (0.30)	6.39 (0.29)
Word Identification Standard Score	110.83 (11.40)	110.71 (11.69)	111.64 (10.77)	111.78 (10.74)
Word Attack Standard Score	112.29 (9.87)	112.76 (9.66)	112.93 (9.32)	113.42 (8.96)
Passage Comprehension Standard Score	98.25 (10.88)	97.92 (10.72)	98.90 (10.76)	98.82 (10.30)
Average Standard Score	107.15 (9.34)	107.06 (9.47)	107.85 (8.96)	107.97 (8.72)

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

the use of this data. Table 6 summarises Lexia use. A median of 31 and inter quartile range (IQR) of 27–33 weeks with Lexia sessions was observed in the intervention students. The median number of total hours using Lexia was 34 (IQR 27 to 42); 269/338 students (79.6 %) completed at least one hour a week for 12 (not necessarily consecutive) weeks (compliance definition one), and 314/338 (92.9 %) completed at least 12 h (compliance definition two).

The CACE analysis results (Table 7) show similar effect sizes to the primary, intention-to-treat analysis providing no evidence that a higher dose of the Lexia program translated to an increased effect in WRMT-III composite reading scores relative to teaching as usual.

Subgroup analysis for low SES students

A total of 265 (38.0 %) students had ever been eligible for FSM (128 intervention, 137 control), of which 233 (87.9 %) had a complete WRMT-III composite reading score (114, 89.0 % intervention, 119, 86.9 % control). In the model that included an interaction between allocation and FSM status, the interaction term was statistically significant at the 10 % significance level (p = 0.053, Table 8). When restricting the primary analysis to only students who had ever been eligible for FSM the adjusted mean difference in WRMT-III composite reading score was 9.47 (95 % CI 0.44–18.52). The Hedges' g effect size is 0.18 in favour of the intervention (95 % CI 0.01 to 0.35), and this difference is statistically significant (p = 0.04).

Implementation fidelity

Analysis of the programme data showed that approximately 80 % of students completed at least one hour of *Core5* usage a week for a minimum of 12 weeks and approximately 93 % completed a minimum of 12 h over a 24-week period. Compliance was therefore deemed to be high

Table 4
Evaluation of effects on primary and secondary outcomes.

	Unadjusted means					Effect size		
	Intervention group		Control group		ICC (95 % CI)			
Outcome	N (missing)	Mean (95 % CI)	N (missing)	Mean (95 % CI)		Total n (intervention; control)	Hedges' g (95 % CI)	p-value
WRMT-III composite score	310 (38)	379.5 (374.0, 385.0)	310 (39)	376.6 (371.3, 382.0)	0.063 (0.010, 0.121)	620 (310; 310)	0.075 (−0.027, 0.176)	0.150
WRMT-III word identification score	318 (30)	95.8 (94.1, 97.5)	322 (27)	94.5 (92.8, 96.2)	0.059 (0.011, 0.117)	640 (318; 322)	0.065 (−0.037, 0.166)	0.213
WRMT-III word attack score	318 (30)	101.1 (99.6, 102.6)	320 (29)	99.8 (98.3, 101.3)	0.039 (0.000, 0.092)	638 (318; 320)	0.098 (−0.029, 0.225)	0.131
WRMT-III passage comprehension score	313 (35)	93.0 (91.7, 94.4)	320 (29)	91.9 (90.5, 93.3)	0.078 (0.024, 0.142)	633 (313; 320)	0.074 (−0.055, 0.203)	0.260
WRMT-III oral reading fluency score	314 (34)	89.0 (87.4, 90.6)	313 (36)	88.6 (87.1, 90.1)	0.042 (0.000, 0.098)	627 (314; 313)	0.047 (−0.061, 0.155)	0.391
KS1 raw reading score	300 (48)	13.8 (12.4, 15.1)	308 (41)	13.8 (12.6, 15.1)	0.109 (0.046, 0.182)	608 (300; 308)	−0.030 (−0.155, 0.096)	0.643

Table 5
Missing data for primary outcomes scale for the targeted sample [Missing data analysis?]

	Adjusted mean difference (95 % CI)	Residual standard error	Hedges' g (95 % CI)	p-value
Multiple imputation by chained equations	3.09 (−1.95, 8.14)	50.53	0.061 (−0.039, 0.161)	0.229
Multiple imputation by chained equations with multilevel model	3.70 (−1.28, 8.68)	50.39	0.073 (−0.025, 0.172)	0.145

Table 6
Summary of compliance in the intervention arm.

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

during the intervention period.

In the teacher-completed surveys, the majority of respondents (92 %, 48 out of 52 schools; teachers from 5 schools did not complete the survey) reported using the reports, although only 63 % (33 schools) said that did so at least weekly. Approximately 1 in 4 schools (27 %, 14 out of 52 schools) reported regularly using the paper-based resources. However, a similar number (23 %, 12 schools) indicated that they did not use the paper-based resources at all. Consequently, implementation of these periphery components of the program was judged to be variable between schools.

Discussion

This was a well conducted study of a computer-assisted reading intervention using the *Core5* software in elementary schools over one academic year with selected struggling readers across 57 schools. The study provides an important contribution to the evidence-base for *Core5* given its design as a larger randomized controlled trial than previously undertaken, randomized at the individual level including analysis for 620 students, across multiple sites in England. Positive, non-significant effect sizes of +0.08 were detected for reading skills for the treatment group who received the intervention in addition to usual teaching of the English national curriculum (2013), compared to the control group who continued with usual teaching. A positive effect size of +0.18 was also detected in improved reading skills for low SES students in the treatment group compared to a control of low SES students (identified using the proxy measure of eligibility for free school meals in school). These findings can be interpreted as 'medium' effect sizes, owing to the large size of the study, the fact that it was situated in the elementary stage, and that it was evaluated using generalized standardized tests (Kraft, 2018; Lipsey et al., 2012). Kraft (2018, 2023) reminds us that it is often difficult to replicate effects of education trials at scale due to implementation challenges, even for program effects with broad external validity. This view is supported by the medical literature which highlights the difficulty of demonstrating an expected effect size when RCTs are scaled in size (Sidebotham & Barlow, 2023). Kraft (2023) uses a dataset of over 3000 effect sizes from randomized controlled trial studies in education which support his earlier findings (Kraft, 2018) that

Table 7
CACE analysis on the primary outcome.

CACE analyses	Correlation[a]	F-test statistic	Adjusted mean difference (95 % CI)	Residual standard error	Hedges' g (95 % CI)	p-value
Completed at least 60 min a week for a minimum of 12 (non-consecutive) weeks	0.80	632	3.27 (−2.95, 9.49)	48.46	0.067 (−0.061, 0.196)	0.304
Completed at least 720 min in total	0.92	1951	2.80 (−2.52, 8.11)	48.42	0.058 (−0.052, 0.167)	0.303
Number of hours of intervention received	0.86	978	0.08 (−0.07, 0.22)	48.44	0.002 (−0.001, 0.005)	0.303

Table 8
Evaluation of effects on primary outcome for FSM subgroup.

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	N (missing)	Mean (95 % CI)	N (missing)	Mean (95 % CI)	Total n (intervention; control)	Hedges' g (95 % CI)	p-value
Ever FSM	114 (14)	373.7 (363.2, 384.1)	119 (18)	366.8 (357.6, 376.1)	233 (114; 119)	0.177 (0.008, 0.345)	0.041

education interventions using standardized achievement outcomes yield lower yet still policy relevant effects. Kraft (2018) suggests that an effect size of 0.05 to less than 0.20 is a 'medium' effect size in education trials. With its robust RCT design, its large sample size, and its 'medium' effect size findings, the current *Core5* study builds on the evidence base of smaller randomized trials in elementary school. For example, O'Callaghan et al. (2016) found similar results in a trial with only 98 children; a small randomized controlled trial of the program with 116 children in five schools found an effect size of +0.23 and was given a 'strong' ESSA rating (Evidence for ESSA, 2024); and Schechter et al. (2015) found a result of ES + 0.53 for low SES students, although this study only included 83 students with randomization at class level. Other larger studies have been quasi-experimental and although not as robust as an RCT design, they also support the benefit of the *Core5* program for improving children's reading (Hobbs, 2016; Wilkes et al., 2020), including for low SES students (Macaruso et al., 2019).

Activity Theory provides an account of learning as a mediated process, by engagement with the activity which the computer supports (Daniels, 2004). In this study, in line with Activity Theory as evolved by Vygotsky (1978) and the 'activity system' adapted by Engeström (2000), the subject (in this case the Grade 1 student) engaged in an activity mediated by a tool (in this case the *Core5* computer assisted program) and was supported within a community (in this case by a teaching assistant who oversaw the learning and intervened as appropriate). The support within the community includes rules and divisions of labour for their functioning (Engeström, 2000). In the case of the *Core5* program intervention delivery, there is a recommended delivery dosage of a minimum of one-hour per week for at least 12 not necessarily consecutive weeks (equivalent to 12 h in total), where students engage with the program. As the naturally occurring data collected in the process evaluation indicated, student engagement with the programme was high. However, in addition to dosage, there is an instruction rule whereby the teaching assistant is tasked with overseeing the intervention delivery and with supporting the student if they are unable to progress within the computer program at any point in the process. Delivery staff therefore had access to downloadable paper resources to support student instruction when required and had access to dashboard-reports with information about student process enabling them to focus their instruction accordingly. Yet delivery staff were not as informed about the additional support which the students required as they might have been with more regular engagement with the dashboard reports or use of the paper-based resources. This 'activity system' provides a helpful lens through which to view the mechanisms of change which gave rise to the findings. As the evidence suggests, we can confirm that the rules of engagement were complied with enabling students to use the *Core5* program with fidelity, in that schools were able to timetable student access to the online program as specified and students engaged with the software for the required time. However, there was extensive variability in teacher's preparedness to support the students to maximise their learning, as approximately half the school staff did not use the structured additional paper-based resources, or informed themselves with the student progress reports regularly, such that they could react swiftly to scaffold learning as it arose.

Given the positive effect sizes detected for all students and for low SES students compared to the control, the lens of the 'activity system' enables us to affirm the potential of using this technology assisted learning. Importantly, the 'activity system' highlighted a problem in the support which students received from the 'community' (teachers) where the rules and divisions of labour for their functioning where not always followed. This meant that staff were not as well informed about student progress as they could have been had they downloaded the progress reports weekly, and students did not always receive the additional instruction they required through the paper-based resources as designed. Although the positive effects on children's reading may have been greater if teachers engaged more fully in downloading student reports and in the use of the program's paper-based resources to support student learning where necessary, the progress in reading made by the students who engaged in the intervention (using the technological 'tool'), compared to a control group supports the potential of using the *Core5* program in elementary schools with struggling readers, to support usual practice. Fidelity of implementation of evidence-informed approaches is accepted as a driver to improvement in the health and education sectors, with acknowledgement that it is difficult to achieve (Aarons, Hurlburt, & Horwitz, 2011; McLeod, Southam-Gerow, Tully, Rodriguez, & Smith, 2013). Using a computer-assisted learning program such as *Core5* provides the conditions required to facilitate usage as per design, as evidenced by the dosage usage during implementation, and the resulting improved outcomes in student reading, although we recognise that improvement is required in encouraging delivery staff to use the resources provided to maximise their scaffolded support to students.

Promoting the ability to read and literacy development is a social entitlement, a key determinant of well-being and a goal of human development (Sen, 1999), enabling children to fully develop their key capabilities necessary to flourish in society (Nussbaum, 2000; Sen, 1992). The ability to learn and develop (including literacy development) are moulded by the transition periods of student's lives from one stage of competence to another. Students exposed to situations where they can develop a competence and are given freedom to exercise it can improve their functioning and form more complex competence sets (Vygotsky, 1978). Those living in literacy environments with greater stimulation, including outside of school, develop and extend their language more readily promoting reading ability and understanding to a greater extent than those who have less exposure to stimulation in this area (Bernstein, 1996). This exposure to words and comprehension facilitates children's ability to reason and understand abstract concepts embedded within texts to which they are exposed in school and is particularly important for children growing up in less stimulated literacy environments where their language may not be stimulated to the same extent as children from more literate backgrounds (Bernstein, 1999). To improve their reading skills the bespoke language exposure for struggling readers engaged in the computer-assisted learning *Core5* program can provide an important context and medium in which to develop the literacy skills for children who struggle, including those who may have had less opportunity to develop these skills previously and who continue to struggle with reading at school.

The findings from this study indicate that coupling the core curriculum with *Core5* improves student reading outcomes for all students when compared to the control group who continued with usual teaching and learning. Improvements for the treatment versus the control group were also found for low SES students who benefitted most from the intervention, with effect sizes of +0.18 detected in improved reading skills. Given that it is recognized that children from lower SES often struggle to reach their full academic potential (Matherly, Amin, & Nahyan, 2017; Pishghadam, 2011) their participation in *Core5* made a positive impact on their reading development which usual teaching aligned to the English national curriculum alone was unable to leverage. In England, socio-economically less advantaged students have on average lower attainment outcomes at school (Gorard, Siddiqui, & See, 2019). For example, in 2019 and 2023, 73 % of all students achieved the expected reading standards at age 11 compared with only 51 % of low SES students in England (Department for Education, 2019a; Department for Education, 2023a, 2023b). This situation has onward effects, as evidenced by only 1.5 % of students with low prior attainment aged 11 meeting the expected standard at age 16 in English and maths (Department for Education, 2019b). In respect of poor academic outcomes, although a lack of literacy experiences outside of school may contribute to this situation, schools from high poverty areas continue to seek means to address inequalities in academic outcomes by improving literacy standards for struggling students. *Core5* clearly demonstrates its potential to improve outcomes for all children struggling with reading, and particularly for children from high poverty backgrounds, thereby contributing to the reduction of inequality in the system.

Limitations

Although the study demonstrated the potential of using the *Core5* program with struggling readers, there were limitations to the study. This trial focused on a small number of struggling readers in Grade 1 per school whereas the programme is designed to span across year groups and more than 6–7 students per year per school. A wider age range would have enabled the researchers to have tested the Logic Model more thoroughly. Nevertheless, the low attrition rates and high levels of compliance do indicate we can be secure in our current findings.

In addition, the limitations of this study are similar to those which Wilkes et al. (2020) found: we also found a key drawback of the present study to be our inability to determine the extent to which various program element usages related to impact. Limitations of our study included lack of data to explain how various elements of the program (particularly teacher support aspects) contributed to reading gains for students in the treatment group. For example, we do not know with precision what teacher support mechanisms were used as this data was collected from perception surveys only, nor do we know how teacher support impacted on children's progress (these included both accessing pupil progress information online on a regular basis, and downloading support paper based resources for use with students who struggle through face-to-face instruction). We also do not know the extent to which levels achieved by students on the software relate to progress made as this detailed data was not collected in the study. It is therefore recommended that future studies focus on this more nuanced detailed data collection to better understand the mechanisms of change which gave rise to the reading progress in participating students.

Conclusion

The results from the RCT reported here provide good evidence that the computer-assisted learning program *Core5* is implementable across a variety of elementary schools in England. Using the program resulted in improved student reading outcomes for the treatment group compared to the control group, measured through an independent standardized reading test not aligned to curriculum materials. In addition to positive, non-significant effect sizes detected for the treatment group, students from high poverty backgrounds in particular were found to have benefited most from the intervention suggesting that *Core5* has the capacity to reduce inequalities in educational outcomes within the English school system.

We therefore recommend that this program be used by schools and be tested in a larger RCT clustered at the school level, to establish whether the effects detected in this study are generalizable more widely across the system.

CRediT authorship contribution statement

Maria Cockerill: Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Louise Tracey:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Louise Elliott:** Project administration, Investigation. **Caroline Fairhurst:** Writing – review & editing, Visualization, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Laura Mandefield:** Writing – review & editing, Visualization, Validation, Methodology, Formal analysis, Data curation. **Imogen Fountain:** Writing – review & editing, Visualization, Validation, Methodology, Formal analysis, Data curation, Conceptualization. **Sarah Ellison:** Formal analysis. **Allen Thurston:** Writing – review & editing, Project administration, Methodology, Funding acquisition. **Joanne O'Keeffe:** Writing – review & editing, Project administration.

Data availability

The authors do not have permission to share data.

References

Aarons, G. A., Hurlburt, M., & Horwitz, S. M. (2011). Advancing a conceptual model of evidence-based practice implementation in public service sectors. *Administration and Policy in Mental Health and Mental Health Services Research, 38*(1), 4–23. https://doi.org/10.1007/s10488-010-0327-7

Al-Huneini, H., Walker, S. A., & Badger, R. (2020). Introducing tablet computers to a rural primary school: An activity theory case study. *Computers & Education, 143*, Article 103648. https://doi.org/10.1016/j.compedu.2019.103648

Azur, M. J., Stuart, E. A., Frangakis, C., & Leaf, P. J. (2011). Multiple imputation by chained equations: What is it and how does it work? *International Journal of Methods in Psychiatric Research, 20*(1), 40–49. https://pubmed.ncbi.nlm.nih.gov/21499542/.

Bernstein, B. (1996). *Pedagogy, symbolic control and identity*. Taylor & Francis.

Bernstein, B. (1999). Vertical and horizontal discourse: An essay. *British Journal of Sociology of Education, 20*(2), 157–173.

Castro, J., & Robnolt, V. (2016). Woodcock Johnson Reading mastery: Test review. *READ, 600*. https://rampages.us/castrojf/wp-content/uploads/sites/23866/2017/04/Reading-Mastery-Paper.docx.

Chambers, B., Slavin, R. E., Madden, N. A., Abrami, P. C., Tucker, B. J., Cheung, A., & Gifford, R. (2008). Technology infusion in success for all: Reading outcomes for first graders. *Elementary School Journal, 109*(1), 1–15.

Chen, J., Wang, M., Kirschner, P. A., & Tsai, C. (2018). The role of collaboration, computer use, learning environments, and supporting strategies in CSCL: A meta-analysis. *Review of Educational Research., 88*(6), 799–843.

Cheung, A., & Slavin, R. E. (June 2012). *(2012). 'Effects of educational technology applications on Reading outcomes for struggling readers: A best evidence synthesis'*. Baltimore, MD: Johns Hopkins University, Center for Research and Reform in Education.

Cornet, V., Voida, S., & Holden, R. J. (2018). Activity theory analysis of heart failure self-care. *Mind, Culture, and Activity, 25*(1), 22–39. https://www.tandfonline.com/doi/full/10.1080/10749039.2017.1372785.

Daniels, H. (2004). Activity theory, discourse and Bernstein. *Educational Review, 56*(2), 121–132. https://www.tandfonline.com/doi/full/10.1080/0031910410001693218.

Department for Education. (2013). English programmes of study: Key stages 1 and 2. *National curriculum in England.* https://assets.publishing.service.gov.uk/media/5a7de93840f0b62305b7f8ee/PRIMARY_national_curriculum_-_English_220714.pdf.

Department for Education (2019a). National curriculum assessments at key stage 2 in England, 2019 (revised). https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/851798/KS2_Revised_publication_text_2019a_v3.pdf (Accessed 02/08/2023).

Department for Education. (2019b). Schools, pupils and their characteristics: January 2019. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/812539/Schools_Pupils_and_their_Characteristics_2019b_Main_Text.pdf Accessed 02/08/2023.

Department for Education. (2022). Key stage 1 and phonics screening check attainment. https://explore-education-statistics.service.gov.uk/find-statistics/key-stage-1-and-phonics-screening-check-attainment Accessed 25/07/2023.

Department for Education. (2023a). Key stage 2 attainment: National headlines. https://explore-education-statistics.service.gov.uk/find-statistics/key-stage-2-attainment-national-headlines/2022-23 Accessed 25/07/2023.

Department for Education. (2023b). The reading framework. https://assets.publishing.service.gov.uk/media/664f600c05e5fe28788fc437/The_reading_framework_.pdf.

Dunn, G., Maracy, M., & Tomenson, B. (2005). Estimating treatment effects from randomized clinical trials with noncompliance and loss to follow-up: The role of

instrumental variable methods. *Statistical Methods in Medical Research, 14*(4), 369–395. https://pubmed.ncbi.nlm.nih.gov/16178138/.
EEF. (2021). Lexia Reading Core5. *Evaluation Protocol (Amended)*. https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/lexia.
Engeström, Y. (1990). *Learning, working, and imagining: Twelve studies in activity theory*. Helsinki, Finland: Orienta-Konsultit Oy.
Engeström, Y. (2000). Activity theory as a framework for analyzing and redesigning work. *Ergonomics, 43*(7), 960–974. https://www.tandfonline.com/doi/abs/10.1080/001401300409143.
Evidence for ESSA. (2024). Lexia Core5 Reading Program – Struggling Readers. https://www.evidenceforessa.org/program/lexia-core5-reading-program-struggling-readers/.
Fairhurst, C., & Rex, S. (2019). *Statistical analysis plan: Lexia*. London: Education Endowment Foundation. https://educationendowmentfoundation.org.uk/public/files/Projects/Lexia_SAP.pdf.
Feister, L. (2013). *Early warning confirmed: A research update on third-grade Reading*. Baltimore, MD: Annie E. Casey Foundation. https://www.aecf.org/resources/early-warning-confirmed.
Ferreira Lemos, M., da Cunha, R., & Júnior, F. (2018). Facebook in Brazilian schools: Mobilizing to fight back. *Mind, Culture, and Activity, 25*(1), 53–67. https://www.tandfonline.com/doi/full/10.1080/10749039.2017.1379823.
Gorard, S., Siddiqui, N., & See, B. H. (2019). The difficulties of judging what difference the pupil premium has made to school intakes and outcomes in England. *Research Papers in Education, 36*(3), 355–379. https://doi.org/10.1080/02671522.2019.1677759
Higgins, S., Xiao, Z., & Katsipataki, M. (2012). *The impact of digital technology on learning: A Summary for the education endowment foundation*. London: Education Endowment Foundation. https://educationendowmentfoundation.org.uk/public/files/Publications/The_Impact_of_Digital_Technologies_on_Learning_(2012).pdf.
Hobbs, J. (2016). Early intervention software program Evaluation: 2015-2016 program results. *Evaluation and Training Institute*. https://www.imaginelearning.com/wp-content/uploads/2022/11/Utah-Early-Intervention-Reading-Software-Program-Program-Evaluation-Results.pdf.
Holen, J. B., Hung, W., & Gourneau, B. (2017). Does one-to-one technology really work: An Evaluation through the Lens of activity theory. *Computers in the Schools., 34*(1–2), 24–44.
Hsieh, Y. H., & Tsai, C. C. (2012). The effect of moderator's facilitative strategies on online synchronous discussions. *Computers in Human Behavior, 28*, 1708–1716.
Hurwitz, L. B., & Macaruso, P. (2021). Supporting struggling middle school readers: Impact of the Lexia® PowerUp Literacy® program. *Journal of Applied Developmental Psychology*, (77), 101329. Elsevier https://doi.org/10.1016/j.appdev.2021.101329.
Hurwitz, L. B., & Vanacore, K. P. (2022). Educational technology in support of elementary students with reading or language-based disabilities: A cluster randomized control trial. *Journal of Learning Disabilities, 56*(6). Sage Journals https://journals.sagepub.com/doi/full/10.1177/00222194221141093.
Issroff, K., & Scanlon, E. (2001). *Case studies revisited: What can activity theory offer? Proceedings of international conference on computer supported collaborative learning 2001 (CSCL 2001), Maastricht, 22–24 march 2001*.
Issroff, K., & Scanlon, E. (2002). Using technology in higher education: An activity theory perspective. *Journal of Computer Assisted Learning, 18*(1), 77–83. ISRCTN42120944 https://doi.org/10.1186/ISRCTN42120944.
Kaptelinin, V., & Nardi, B. (2012). *Activity theory in HCI: Fundamentals and reflections*. San Rafael, California: Morgan and Claypool.
Kaptelinin, V., & Nardi, B. (2018). Activity theory as a framework for human-technology interaction research. *Mind, Culture, and Activity., 25*(1), 3–5. https://doi.org/10.1080/10749039.2017.1393089
Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman test of educational achievement* (2nd ed.). San Antonio, TX: NCS Pearson.
Kraft, M. A. (2018). *Interpreting effect sizes of education interventions*. Brown University Working Paper. https://scholar.harvard.edu/files/mkraft/files/kraft_2018_interpreting_effect_sizes.pdf.
Kraft, M. A. (2023). The effect-size benchmark that matters most: Education interventions often fail. *Educational Researcher., 52*(3), 131–187. https://doi.org/10.3102/0013189X231155154
Lewin, C., Smith, A., Morris, S., & Craig, E. (2019). Using digital technology to improve learning. *Evidence Review. Education Endowment Foundation.* https://d2tic4wvo1iusb.cloudfront.net/production/documents/guidance/Using_Digital_Technology_to_Improve_learning_Evidence_Review.pdf?v=1717064802.
Liao, Y. C., & Hao, Y. (2008). Large -scale studies and quantitative methods. In J. Voogt, & G. Knezek (Eds.), *International handbook of information Technology in Primary and Secondary Education* (pp. 1019–1035). Springer. https://teachwithict.wordpress.com/wp-content/uploads/2011/08/dede.pdf.
Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., … Busick, D. (2012). *Translating the effects of education interventions into more readily interpretable forms*. National Center for Special Education Research. Institute of Education Sciences. https://ies.ed.gov/ncser/pubs/20133000/pdf/20133000.pdf.
Macaruso, P., Wilkes, S., Franzén, S., & Schechter, R. (2019). Three-year Longtitudinal study: Impact of a blended learning program – Lexia Core5 Reading – On Reading gains in low-SES kindergarteners. *Computers in the Schools, 36*(1), 2–18. https://doi.org/10.1080/07380569.2018.1558884
Matherly, L. L., Amin, N., & Nahyan, A. S. K. A. (2017). The impact of generation and socioeconomic status on the value of higher education in the UAS: A longitudinal study. *International Journal of Educational Development, 55*, 1–10.
McAvinia, C. (2016). *Activity theory. Online learning and its users, 59–100*. Elsevier Ltd. (ISBN 978-0-08-100626-9.).
McLeod, B. D., Southam-Gerow, M. A., Tully, C. B., Rodriguez, A., & Smith, M. M. (2013). Making a case for treatment integrity as a psychosocial treatment quality indicator for youth mental health care. *Clinical Psychology: Science and Practice, 20*(1), 14–32. https://doi.org/10.1111/cpsp.12020
McMurray, S. (2012). An evaluation of the use of Lexia Reading software with children in year 3, Northern Ireland (6- to 7-year olds). *Journal of Research in Special Educational Needs., 13*(1), 15–25. Nasen https://nasenjournals.onlinelibrary.wiley.com/doi/epdf/10.1111/j.1471-3802.2012.01238.x?saml_referrer.
McNally, S., Ruiz-Valenzuela, J., & Rolfe, H. (2018). *ABRA: Online Reading Support Evaluation Report and Executive Summary. October 2016. Addendum added march 2018*. London: Education Endowment Foundation. https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/ABRA_with_addendum.pdf?v=1717062350.
Michinov, N., & Primois, C. (2005). Improving productivity and creativity in online groups through social comparison process: New evidence for asynchronous electronic brainstorming. *Computers in Human Behavior, 21*, 11–28.
National Assessment of Educational Progress (NAEP). (2022). NAEP report card: Reading. https://www.nationsreportcard.gov/reading/nation/achievement/?grade=8.
National Institute of Child Health and Human Development. (2000). *Report of the national reading panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office
Nussbaum, M. C. (2000). *Women and human development: The capabilities approach*. Cambridge University.
O'Callaghan, P., McIvor, A., McVeigh, C., & Rushe, T. (2016). A randomised controlled trial of an early- intervention, computer-based literacy program to boost phonological skills in 4- to 6-year old children. *British Journal of Educational Psychology, 86*(4), 546–558. https://pubmed.ncbi.nlm.nih.gov/27473924/.
Pishghadam, R. (2011). Parental education and social and cultural capital in academic achievement. *Inter- national Journal of English Linguistics, 1*(2), 50–57.
Quartagno, M., & Carpenter, J. (2016). Jomo: Multilevel joint modelling multiple imputation [data collection]. *The Comprehensive R Archive Network*. https://datacompass.lshtm.ac.uk/id/eprint/397/.
R Development Core Team. (2011). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. http://www.R-project.org.
Schechter, R., Macaruso, P., Kazakoff, E. R., & Brooke, E. (2015). Exploration of a blended learning approach to reading instruction for low SES students in early elementary grades. *Computers in the Schools, 32*(3–4), 183–200. https://www.tandfonline.com/doi/full/10.1080/07380569.2015.1100652.
Sen, A. (1992). *Inequality re-examined*. Oxford University Press.
Sen, A. (1999). *Development as freedom*. Alfred Knopf.
Sidebotham, D., & Barlow, C. J. (2023). The winner's curse: Why large effect sizes in discovery trials always get smaller and often disappear completely. *Anaesthesia, 79*(1), 86–90. Association of Anaesthetists, John Wiley & Sons Ltd https://doi.org/10.1111/anae.16161.
Slavin, R. E., Lake, C., Chambers, B., Cheung, A., & Davis, S. (2009). Effective Reading programs for the elementary grades: A best-evidence synthesis. *Review of Educational Research, 79*(4), 1391–1466. http://www.jstor.org/stable/40469101.
Standards & Testing Agency. (2020). Appendix C: 2019 validity framework. Key stage 1 English reading. January 2020 https://assets.publishing.service.gov.uk/media/5e56585c86650c10ef4d3cf8/Appendix_C_2019_validity_framework_key_stage_1_English_reading.pdf.
StataCorp. (2017). *Stata statistical software: Release 15*. College Station, TX: StataCorp LLC. https://academic.oup.com/ej/article-abstract/102/415/1581/5157164.
Torgerson, C. J., Brooks, G., & Hall, J. (2006). *A systematic review of the research literature on the use of phonics in the teaching of Reading and spelling*. DFES Publications. pp83 https://www.researchgate.net/publication/265619755_A_Systematic_Review_of_the_Research_Literature_on_the_Use_of_Phonics_in_the_Teaching_of_Reading_and_Spelling.
Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2010). Computer assisted instruction to prevent early reading difficulties in students at risk for dyslexia: Outcomes from two instructional approaches. *Ann Dyslexia, 60*(1), 40–56. Epub 2010 Jan 6. PMID: 20052566; PMCID: PMC2888606 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2888606/.
Tracey, L., Chambers, B., Slavin, R., Hanley, P., & Cheung, A. (2014). Success for all in England: Results from the third year of a National Evaluation. *SAGE Open, 4*(3).
Tracey, L., Elliott, L., Fairhurst, C. M., Mandefield, L., Fountain, I., & Ellison, E. (2022). *Lexia Core Reading: Evaluation Report*. Education Endowment Foundation. https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/lexia.
Vogel, F., Wecker, C., Kollar, I., & Fischer, F. (2017). Socio-cognitive scaffolding with computer-supported collaboration scripts: A meta-analysis. *Educational Psychology Review, 29*, 477–511.
Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
Wainer, J., Dwyer, T., Dutra, R. S., Covic, A., Magalhães, V. B., Ferreira, L. R., … Claudio, K. (2008). Too much computer and Internet use is bad for your grades, especially if you are young and poor: Results from the 2001 Brazilian SAEB. *Computers and Education, 51*, 1417–1429. https://eric.ed.gov/?id=EJ807640.
Wechsler, D. (2009). *The Wechsler individual achievement test* (3rd ed.). San Antonio, TX: NCS Pearson.
Wilkes, S., Kazakoff, R., Prescot, J. E., Bundschuh, K., Hook, P. E., Wolf, R., … Macaruso, P. (2020). Measuring the impact of a blended learning model on early literacy growth. In *Journal of computer assisted learning, 36(5), 581–779*. Wiley Online Library. https://onlinelibrary.wiley.com/doi/epdf/10.1111/jcal.12429.

Woll, A., & Bratteteig, T. (2018). Activity theory as a framework to analyze technology-mediated elderly care. *Mind, Culture, and Activity, 25*(1), 6–21. https://www.tandfonline.com/doi/abs/10.1080/10749039.2017.1375528.

Woodcock, R. W. (1998). *Woodcock Reading mastery tests – Revised/normative update*. Minneapolis, MN: Pearson Assessments.

Woodcock, R. W. (2011). *Woodcock Reading mastery tests* (3rd ed.). Bloomington, MN: NCS Pearson.

World Bank. (2022). The State of Global Learning Poverty: 2022 Update. https://thedocs.worldbank.org/en/doc/e52f55322528903b27f1b7e61238e416-0200022022/original/Learning-poverty-report-2022-06-21-final-V7-0-conferenceEdition.pdf.

Zheng, X. L., Kim, H. S., Lai, W. H., & Hwang, G. J. (2019). Cognitive regulations in ICT-supported flipped classroom interactions: An activity theory perspective. *British Journal of Educational Technology, 51*(1), 103–130. https://doi.org/10.1111/bjet.12763