

## **Case Study 1: An Evidence-Based Practice Review Report**

**Theme: School/setting Based Interventions for Learning.**

***How effective is Lexia® Core5® Reading at improving the reading skills of primary-aged children?***

### **Summary**

Lexia® Core5® Reading ([www.lexiauk.co.uk](http://www.lexiauk.co.uk)) is an online reading intervention produced by Lexia Learning; it offers its users a personalised learning pathway through an initial auto-placement task, and subsequent scaffolded tasks designed to support and challenge the learner. Research demonstrating its efficacy, and that of a previous version of the intervention (Lexia Reading) is mainly US based with studies evaluating its effectiveness as a component of a blending learning approach alongside traditional teacher-led instruction. The current review will examine the effectiveness of the most recent version of the intervention (Lexia® Core5® Reading) for improving reading skills in primary-aged children by evaluating this research. A systematic review of the literature identified six studies which met the inclusion criteria; these were then evaluated using the Harden and Gough (2012) Weight of Evidence Framework, with the Gersten et al. (2005) coding protocol used to assess methodological quality. Significant results and medium-large effect sizes were found in the majority of the studies, but lack of randomised control trials limit knowledge about the intervention's effectiveness. Future studies need to address the question of effectiveness through use of comparison and control groups to remove potentially confounding variables and assess whether the chances of intervention success are likely for all groups of children. Studies that include a qualitative element are also recommended to understand teacher and student motivation which could have an impact on implementation success.

## Introduction

### *Intervention*

Lexia® Core5® Reading (Lexia) is an online computer program suitable for all primary-aged children from Reception to Year 6. It can be used as an intervention at the targeted, group or universal level to improve children's reading. The cost for purchasing the program is dependent on the number of licenses that are purchased but can range from £18-£25 per student/per year. Lexia provides differentiated instruction in six areas of reading: phonological awareness, phonics, structural analysis, automaticity/fluency, vocabulary and comprehension. It aims to engage and motivate its users through the use of auditory and visual feedback, songs, video and humour alongside printable certificates. The central premise for success is the ability of the program to personalise learning pathways for its users through the use of adaptive technology; the program places the child at the correct level for their learning through an initial auto placement test and then adapts to their strengths and needs through scaffolded activities inherent in each unit of learning. Each unit is adaptable and contains three steps: *standard*, *guided practice* and *instruction*. The *standard step* is the most complex and must be completed to progress to the next unit. If the child makes errors at this step, they are directed to the *guided practice* step where scaffolded practice is provided. There is further support provided in the *instruction step* (if errors are still present), where the skill is explicitly taught and more scaffolded practice given.

Lexia also aims to support teachers by providing access to online data reports on each child; this can be used to track progress and program usage. In addition, paper-based resources are provided which can be used in the classroom to reinforce learning taking place online or used as homework.

For optimum success, Lexia advises three practice goals:

1. Children need to be accessing the program for 3 x 20-30 mins/week

2. Teachers need to monitor children's progress reports (typically once/week)
3. If children are not making progress, children need to receive additional support through paper-based resources linked to their performance in Lexia.

In addition, Lexia provides a clear support system for schools using the program to ensure fidelity of implementation. All support is online and includes initial launch training such as data coaching and technical set up by a Lexia consultant. This is then supplemented by further support by the Lexia team through technical support and live or recorded webinar training.

Studies of effectiveness of a previous version of the program (Lexia Reading) are available (McMurray, 2013; Macaruso & Rodman, 2011; Macaruso & Walker, 2008) but Lexia® Core5® is the current and revised program available to purchase in the UK and as such is the primary focus of this review. The current form of the program is an enhanced version of Lexia Reading with enhancements to audio, visual and animations, alongside a more sophisticated auto-placement system, and easier access to tracked targets; it is also the only version now available to purchase.

### *Psychological Basis*

The Simple View of Reading (Gough & Tunmer, 1986) puts forward the notion that reading comprehension is the product of linguistic comprehension and decoding skills. This conceptual framework was embedded in the Rose Report (2006) which led to systematic phonics instruction becoming the favoured method of early reading instruction in the UK, and the subsequent advent of phonics programs that focused on areas believed to underpin these central skills; phonemic awareness (Melby-Lervag, Lyster, & Hulme, 2012), explicit phonics instruction (Duff et al., 2014), fluency and vocabulary (Melby-Lervag & Lervag, 2014).

Technology-based interventions designed to support these skills are in increasing use with some inherent advantages including embedded scaffolding and an ability to adapt to individual performance (O'Brien, Habib, & Onnis, 2019). Scaffolding (Wood, Bruner, & Ross, 1976)

refers to the social guidance and interaction which takes place within a child's Zone of Proximal Development (Vygotsky, 1978) that builds on existing knowledge to achieve maximum learning potential; in this scenario, the computer provides the guidance and interaction. Research is unclear about whether a computer alone can provide the necessary social aspects of learning (Cheung & Slavin, 2013), and whether teachers need to scaffold the learning that takes place (Mercer & Fisher, 1997). Interventions like Lexia that consider the importance of online and offline teacher-supported resources are therefore best placed to ensure maximum learning potential.

### *Rationale and Relevance*

Raising literacy levels is a central concern for educational psychologists (MacKay, 2002). However, many children are still leaving primary school without the expected levels in reading and writing, with the chances doubled for a child from a lower socioeconomic background (Higgins et al., 2017). This inequality is even starker when assessed alongside the known negative outcomes for adults with low literacy skills which include unemployment, poor health and lower life expectancy (Gilbert, Teravainen, Clark, & Shaw, 2018). Recent guidance from the Education Endowment Fund on improving literacy in Key Stage 1 (Higgins et al., 2017) recommends implementation of a systematic phonics programme, the use of high-quality information about pupils' capabilities and high-quality structured interventions to support pupils. One-to-one or small group tutoring which can be adaptive to a child's needs is also seen a key feature of a successful literacy curriculum (Slavin, Lake, Davis, & Madden, 2011). However, the reality of school funding cuts mean that 83% of schools will be worse off in 2020 than in 2015 (Baisley, 2019), with the affordability of small group and personalised support becoming less achievable. Schools and educational psychologists are therefore keen to explore cost-effective technology-based interventions to support literacy (Picton, 2019). Lexia is one such intervention that has a growing user base in the UK (3000+ schools); it meets many of the recommendations in the EEF guidance in its personalised and adaptive approach.

The current review aims to explore the effectiveness of Lexia in improving reading skills for primary-aged children.

*Review question*

How effective is Lexia® Core5® in improving reading skills for primary-aged children?

**Critical Review of the Evidence Base**

*Literature search*

On 16<sup>th</sup> December 2019, a literature search was conducted using four online electronic databases: ERIC, Web of Science, PsychInfo and UCL Explore. The search terms used were “lexia” AND reading AND intervention. Papers were identified (N=61) and imported into a database (Mendeley) with 18 papers excluded due to duplication; all papers were then screened according to the exclusion criteria (Table 1). This led to the exclusion of 18 papers with 25 papers then screened for titles and abstracts. A final full text screening of 14 papers was conducted with 6 studies found to be eligible for inclusion in the study (see Figure 1 for a flow diagram of the screening process; see Table 2 for final study selection; see also Appendices A and B for more detailed information about included studies).

Table 1

*Inclusion and Exclusion Criteria*

Study Feature	Inclusion Criteria	Exclusion Criteria	Rationale
1. Language	Studies published in English	Studies published in other languages	The reviewer is only fluent in English.
2. Intervention	Studies that use Lexia® Core5®	Studies that do not use Lexia® Core5®	This review is only focused on the impact of Lexia® Core5®, and not earlier iterations of the program which are no longer available to purchase in the UK (Lexia Reading). This is to ensure that the findings are relevant to schools who are using the findings to influence a purchase decision. It also does not look at a program intended for use with older children (PowerUp Literacy™) as is beyond the scope of the review.
3. Outcomes	At least one of the primary outcomes to be related to children’s reading ability	None of the primary outcomes being related to reading ability	Lexia® Core5® is a program that claims to improve children’s reading skills so the review is looking at these specific skills
4. Type of publication	Peer-reviewed journals	Grey literature including non-peer reviewed journals and dissertations	To ensure that research is of a high standard, and has undergone scrutiny by others in the same field

Study Feature	Inclusion Criteria	Exclusion Criteria	Rationale
5. Participants	Primary school-aged children	Children who are above or below primary school age (>4 or <12)	This review looks at the impact of Lexia® Core5® which is designed to be used with primary school-aged children
6. Study Design	Study collects pre- and post- measures of intervention impact	Study does not collect any pre- and post- measures of intervention impact; meta-analyses or systematic reviews	Review was interested in the effectiveness of the intervention which would need to be measured by measures of reading skill before and after children have received the Lexia intervention

Figure 1

Flow Diagram of Selection and Screening Process

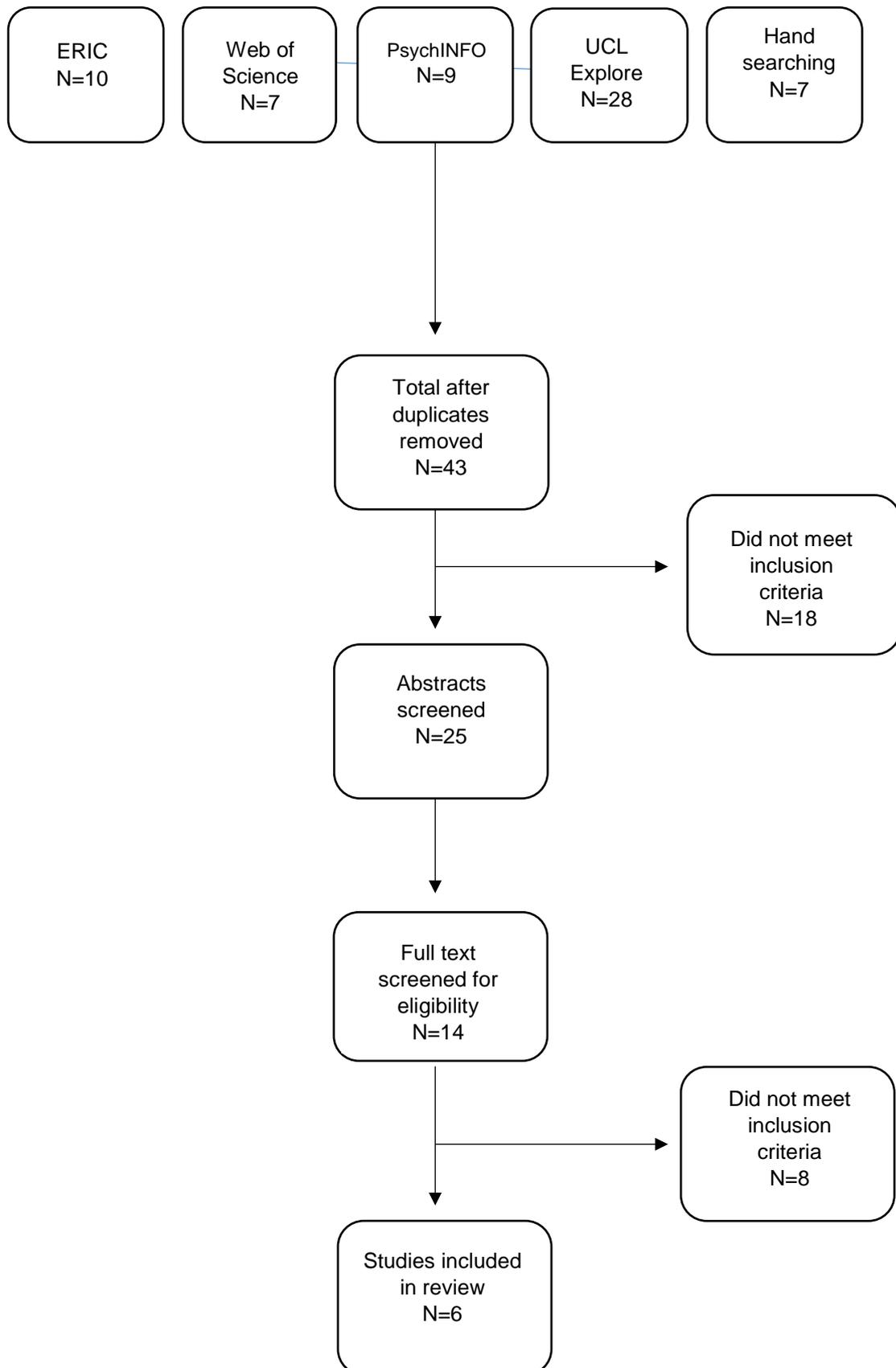


Table 2

---

*The Final Six Studies Included in the Systematic Review*

---

Baron, L., Hogan, T., Schechter, R., Hook, P., & Brooke, E. (2019). Can educational technology effectively differentiate instruction for reader profiles? *Reading and Writing*, 32(9), 2327-2352.

Kazakoff, E., Macaruso, R., & Hook, P. (2018). Efficacy of a blended learning approach to elementary school reading instruction for students who are English Learners. *Educational Technology Research and Development*, 66(2), 429-449.

Macaruso, P., Wilkes, S., Franzén, S., & Schechter, R. (2019). Three-Year Longitudinal Study: Impact of a Blended Learning Program-Lexia® Core5® Reading-on Reading Gains in Low-SES Kindergarteners. *Computers in the Schools*, 36(1), 2-18.

O'Callaghan, P., McIvor, A., McVeigh, C., & 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. *The British Journal of Educational Psychology*, 86(4), 546-558.

Prescott, J., Bundschuh, K., Kazakoff, E., & Macaruso, P. (2018). Elementary school-wide implementation of a blended learning program for reading intervention. *The Journal of Educational Research*, 111(4), 497-506.

Schechter, R., Macaruso, P., Kazakoff, E., & 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.

---

### *Weight of Evidence*

The final six studies were evaluated according to the Weight of Evidence (WoE) framework (Harden and Gough, 2012) which outlines three dimensions to judge the studies against:

- i. Weight of Evidence A (WoE A) which evaluates methodological soundness; this was assessed using an adapted version of the Gersten et al. (2005) framework. Full details and rationale for adaptations can be found in appendices C and D
- ii. Weight of Evidence B (WoE B) which considers methodological relevance in relation to the review question;
- iii. Weight of Evidence C (WoE C) which assesses the relevance of the study to the review question.

Scores for each dimension were then averaged to produce a final overall weight of evidence score (WoE D – see Table 3). Full rationales for judgements based on each of the dimensions can be found in Appendix C.

Table 3

*Overall Weight of Evidence Ratings*

Studies	WoE A	WoE B	WoE C	WoE D
O' Callaghan et al. (2016)	2	3	3	2.67 (High)
Baron et al. (2019)	1	1	1	1.00 (Low)
Schechter et al. (2015)	3	3	2	2.67 (High)
Prescott et al. (2018)	2	1	2	1.67 (Low)
Macaruso et al. (2019)	2	1	2	1.67 (Low)
Kazakoff et al. (2018)	1	1	1	1.0 (Low)

*Note.* Final Weight of Evidence scores were translated into an overall judgement rating based on a score range: Low (1-1.6); Medium (1.7-2.3); High (2.4-3).

*Critical Review of Included Studies*

*Participants*

Participants were all enrolled in primary school settings (N = 2343): their age ranged from 4 – 11 years. Power analysis was carried out by two studies (Kazakoff, Macaruso, & Hook, 2017; O'Callaghan, Mclvor, McVeigh, & Rushe, 2016) with both studies using appropriate sample sizes to detect an effect size at a 0.5 significance level (Cohen, 1988).

Sample sizes ranged from 98 – 884, with three out of the six studies taking place in single schools (Macaruso, Wilkes, Franzen, & Schechter, 2019; Prescott, Bundusch, Kazakoff, & Macaruso, 2018; Schechter, Macaruso, Kazakoff, & Brooke, 2015), one study taking part in two schools (O'Callaghan et al., 2016) and two studies using the same participant data set sourced from 64 schools across one US state (Baron, Hogan, Schechter, Hook, & Brooke,

2019; Kazakoff et al., 2017). Detailed participant demographics were included in all papers except Baron et al. (2019) which impacted on its WoE A rating as it limits the generalisability of the findings. No studies reported any effect of gender on overall findings. The number of participants receiving free school meals was reported in five out of the six studies ( $M=75.4%$ , UK mean in primary schools, 2019 = 15.8%). This was notable considering the increased likelihood of children from a lower SES background leaving school without the required literacy skills (Higgins et al., 2017), and the benefits of an intervention that would work with this population.

Five of the studies were carried out in the US (Baron et al., 2019; Macaruso et al., 2019; Prescott et al., 2018 and Schechter et al., 2015) which has a different educational context and curriculum than the UK. As this could impact on the application of the studies' findings to UK schools, this contributed to their lower WoE C compared to the one study carried out in the UK (O'Callaghan et al., 2016). Participants were drawn from universal school populations except one study (Kazakoff et al., 2017) who used a sample of students who had English as an Additional Language (EAL); this was less relevant to the review question which impacted on its score in WoE C.

Attrition rates were discussed in two of the studies (Macaruso et al., 2019 – 6% over 3 years; O'Callaghan et al., 2016 – 9%, although their data was still included in the final analysis). O'Callaghan et al. (2016) provides anecdotal feedback suggesting that this was caused by frustrations with the technology and pupil absence whilst Macaruso et al., (2019) state that attrition was caused by children leaving school or not being present for assessments. One of the studies had no attrition (Schechter et al., 2015), whilst one used an existing data set so did not report it (Baron et al., 2019). The final two studies (Kazakoff et al., 2017 – 85% over 2 years; Prescott et al., 2018 – 16%;) had attrition but did not provide explanation. Schulz and Grimes (2002) report that an attrition rate greater than 20% poses a serious threat to validity, even with relatively equal drop-out rates for both groups; this potential bias was reflected in a

lower WoE A score for Kazakoff et al., 2017. The relationship between technological difficulties and attrition rates would need to be further explored to ensure that they are not a barrier to the effectiveness of the intervention.

### *Study design*

Two of the six studies used a randomised control trial (O'Callaghan et al., 2016; Schechter et al., 2015) with one study (O'Callaghan et al., 2016) using a wait-list control group. This resulted in high ratings for WoE B and WoE D as the study design provides the highest internal validity and is better placed to answer questions about effectiveness (Petticrew & Roberts, 2006). O'Callaghan et al. (2016) also used both post and 2-month follow-up tests to ascertain whether intervention effects were maintained resulting in a higher rating for WoE B than the Schechter et al. (2015) study.

One study used a longitudinal study design with no control group (Macaruso et al., 2019). Three studies (Baron et al., 2019; Kazakoff et al., 2017; Prescott et al., 2018) used a pretest-posttest design also without control groups resulting in low WoE B scores for all due to difficulties in reliably teasing apart the positive effects of Lexia from other confounding factors such as school effect in studies with more than one school (Kazakoff et al., 2017), or concurrent literacy interventions as reported in two of the studies (Prescott et al., 2018; Macaruso et al., 2019).

### *Measures*

All studies used standardised measures of reading skill (Appendix A) although these differed between studies. O'Callaghan et al. (2016) used four subtests (Blending, Non-Word reading, Phonological Working Memory and Phoneme Segmentation) from the Phonological Assessment Battery 2<sup>nd</sup> Edition (PhAB-2; Gibbs & Bodman, 2014). Two studies (Baron et al., 2019; Kazakoff et al., 2017) used aimsweb (Pearson, 2012) and three studies (Macaruso et al., 2019; Prescott et al., 2018; Schechter et al., 2015) used the Group Reading Assessment

and Diagnostic Evaluation (GRADE; Williams, 2001). One study (Baron et al., 2019) also included performance on Lexia as an outcome measure. The construct validity of the chosen measures was discussed in two of the studies which was reflected in their WoE A rating (Baron et al., 2019; O'Callaghan et al. 2016).

### *Intervention*

All studies used Lexia as per the inclusion criteria, however only one study (O'Callaghan et al., 2016) was focused on analysing the effectiveness of the intervention itself, rather than as a component of a blended learning program; this focus led to a double weighting in WoE C as was the most pertinent to the review question. To assess external validity and to gain a reliable picture of the effectiveness of Lexia, this review assessed the extent that each study implemented the intervention with fidelity. Using the Lexia guidelines for best practice, each study's fidelity of implementation was reflected in their WoE A and C scores. Three of the studies (Macaruso et al., 2019; Prescott et al., 2018; Schechter et al., 2015) all reported that their research teams had conducted observations to assess learner engagement; all studies reported inter-observer agreement at over 90%, and that student engagement ranged from 84-97%. Participants in all studies met minimum weekly usage requirements (20-60 mins/per week) and five out the six studies maintained the intervention for more than the specified amount of time recommended by Lexia (20 weeks). A decline in usage over time (Schechter et al., 2015) and by age (Prescott et al., 2018) was reported which could be due to a decrease in motivation, or other contextual issues. The one study that met weekly usage requirements but not the overall recommended time scale (O'Callaghan et al., 2016) carried out the intervention for 8 weeks.

Lexia recommends using the additional paper-based resources for children who require extra support. This is an important element as these need to be delivered with adult-mediated support and can be used in conjunction with parents. O'Callaghan et al. (2016) was the only study to *not* give children access to this support due to timetabling limitations; it is unclear

what the effect on efficacy of the intervention would be with the addition of adult-led support, and this omission decreases the external validity of its findings. However, it still scored highly on all dimensions due to the rigour of its study design, and the fact that the study does recognise the omission as a limitation.

In addition, two of the studies (Macaruso et al., 2019; Schechter et al., 2015) conducted teacher interviews to learn more about implementation which highlighted how the additional paper-based resources were used, and how 'mini-celebrations' were held for students who had completed a level (Macaruso et al., 2019); this qualitative evidence is a useful indicator of successful implementation, and could be explored in future studies to ascertain what aspects of Lexia are most engaging and successful for children.

#### *Outcomes and Effect Sizes*

Effect sizes were reported in all studies but as they reported different measures of effect size (partial eta squared; Cohen's  $d$ ; Cohen's  $f$ ), all were converted to Cohen's  $d$  using an effect size calculator ([www.psychometrica.de/effect\\_size.html](http://www.psychometrica.de/effect_size.html)) for ease of comparison (Table 4). Average effect sizes were reported for Baron et al. (2019) as the review was interested in the effect of Lexia for a universal classroom population, rather than any specific reader type so the average indicated the effect that could be detected overall in a typical classroom which would consist of all the reader groups listed in the study. However, it is important to note that averaging the effects affects generalisability as it conflates the effect of the intervention for some groups and underestimates it for other groups. A future review could focus on the effect of Lexia for specific populations to ensure generalisability of findings. In accordance with Cohen (1988), effect sizes in this review are interpreted as small (0.2), medium (0.5) and large (0.8). Medium to large effect sizes were found in the majority of studies.

Four out of the six studies (Baron et al., 2019; Kazakoff et al., 2018; Macaruso et al., 2019; Prescott et al. 2018) use a within-subjects study design which could explain the large effect

sizes identified. These effects could be inflated due to the increased power of a within-subjects design which could be seen as a benefit of the design, which is able to make a distinction between whether variability is the result of individual difference or intervention effect (Lakens, 2013). However, this has also been interpreted as a negative which results in an overestimation of the true effect size (Dunlap, Cortina, Vaslow, & Burke, 1996). For the current review, these effect sizes serve as a useful comparative tool between studies with similar designs but are treated with caution when compared to the between-subjects study designs (O'Callaghan et al., 2016; Schechter et al., 2015). Findings from one study (Prescott et al., 2018) suggest that intervention effects decrease with age (K,  $d= 2.78$  – 5<sup>th</sup> Grade,  $d=0.88$ ). This could be due to Lexia having more effectiveness as an early intervention or could be explained through the importance of adhering to usage requirements as outlined above.

Further note needs to be taken when assessing the implications of these effect sizes as the four aforementioned studies do not have a control group, so it is impossible to ascertain whether these studies outcomes can be derived from the impact of Lexia; this limitation is reflected in lower WoE B and overall WoE D scores for all these studies. Schechter et al. (2015) found that the control group who took part in mainstream literacy lessons following the LEAD21 (Wright Group, n.d) literacy curriculum still made statistically significant gains over time ( $d=0.88$ ) which again further questions the intervention impact in the studies without a control group, as the identified gains in the four studies without control groups could be down to many factors present in a successful literacy curriculum.

The two RCTs (O'Callaghan et al., 2016; Schechter et al., 2015) both report similar medium effect sizes (0.52 – 0.55) for blending, non-word reading and comprehension, and small effects for phoneme segmentation (0.38) and vocabulary (0.09). A regression analysis conducted by O'Callaghan et al. (2016) suggests that phonological working memory predicted the impact that the Lexia would have ( $p=.001$ ) which has important implications for considering whether

the intervention will be effective for children with deficits in this area. Baron et al.'s (2019) results also indicate that intervention success is dependent on specific abilities.

Table 4

Effect Sizes for Overall Weight of Evidence Ratings for Included Studies

Study	Measure	Comparison	Effect Size	Descriptor	WoE D
O' Callaghan et al. (2016) N=98	Phonological Assessment Battery 2 <sup>nd</sup> Edition (PhAB-2; Gibbs & Bodman, 2014):	Lexia intervention vs. waitlist control	$d = 0.52$	Medium	High
	<b>Blending</b>				
	Phonological Assessment Battery 2 <sup>nd</sup> Edition (PhAB-2; Gibbs & Bodman, 2014):		$d = 0.55$	Medium	
	<b>Non-word reading</b>				
	Phonological Assessment Battery 2 <sup>nd</sup> Edition (PhAB-2; Gibbs & Bodman, 2014):		$d = 0.38$	Small	
	<b>Phoneme segmentation</b>				
Baron et al. (2019) N=594	aimsweb (Pearson, 2012):	Reader profiles	Poor comprehenders ( $d=0.18$ )	Large (average score)	Low
	<b>Oral Reading Fluency (R-CBM)</b>		Mixed deficits ( $d=1.03$ )		
			Poor decoders ( $d=1.79$ )		
			Typical readers ( $d=0.43$ )		
	Average $d = 0.86$				
	aimsweb (Pearson, 2012):		Poor comprehenders ( $d = 2.68$ )	Large (average score)	
<b>Reading Comprehension (Maze)</b>		Mixed deficits ( $d = 0.94$ )			
		Poor decoders ( $d=0.82$ )			
		Typical readers ( $d=0.17$ )			
		Average $d=1.15$			
	Lexia Core5 Performance		$d = 0.49$	Small	
Kazakoff et al. (2017)	aimsweb (Pearson, 2012): <b>Oral</b>		K grade: $d = 3.02$	Large	Low

N=884	<b>Reading Fluency</b> (R-CBM)		1 <sup>st</sup> grade: $d = 2.82$		
			2 <sup>nd</sup> - 5 <sup>th</sup> grade: $d = 1.76$		
Study	Measure	Comparison	Effect Size	Descriptor	WoE D
Schechter et al. (2015) N=83	Group Reading Assessment and Diagnostic Evaluation (GRADE; Williams, 2001): <b>Vocabulary</b>	Blended learning vs non-blended learning (Lexia as online component)	$d = .09$	Small	High
	Group Reading Assessment and Diagnostic Evaluation (GRADE; Williams, 2001): <b>Comprehension</b>		$d = .52$	Medium	
Prescott et al. (2018) N=641	Group Reading Assessment and Diagnostic Evaluation (GRADE; Williams, 2001): <b>Composite score</b>	Blended learning vs non-blended learning (Lexia as online component)	K grade: $d = 2.78$  1 <sup>st</sup> grade: $d = 1.15$  2 <sup>nd</sup> grade: $d = 1.31$  3 <sup>rd</sup> grade: $d = 0.92$  4 <sup>th</sup> grade: $d = 0.13$  5 <sup>th</sup> grade: $d = 0.88$	All large except 4 <sup>th</sup> grade	Low
Macaruso et al. (2019) N=83	Group Reading Assessment and Diagnostic Evaluation (GRADE; Williams, 2001): <b>Composite score</b>	No comparison	All students: $d = 1.87$  Struggling readers: $d = 2.24$	Large	Low

Note. Effect size descriptors for  $d$  (Cohen, 1988): small,  $d = 0.2$ ; medium,  $d = 0.5$ ; large,  $d = 0.8$ ; K (Kindergarten)

## **Conclusions and Recommendations**

The purpose of this review was to evaluate the effectiveness of Lexia on improving children's reading skill. In line with claims on the Lexia website, the studies that were included in this review do show that the intervention has a positive impact on reading skills as demonstrated by statistically significant results and mainly medium-high effect sizes. However, this review found that the majority of the studies did not use a randomised control trial, which limited answers around effectiveness, and which contributed to four out of the six studies achieving a low overall weight of evidence assessment. Caution needs to be applied to interpretation of findings in these studies due to validity issues that can be inherent within the one-group pretest-posttest design (Barker, Pistrang, & Elliot, 2016). So although the results from this review are promising, the educational significance is still limited. Further experimental research needs to be carried out with control or comparison groups to rule out potentially confounding variables (such as school effect or concurrent literacy interventions) and also examine which specific elements of the intervention are likely to be most responsible for reading improvements. This would further reinforce the positive findings from the O'Callaghan et al. (2016) and Schechter et al. (2015) which received high overall weight of evidence scorings due to their methodological soundness and relevance to the review question.

Moreover, questions arise from the review about the importance of fidelity of implementation and inclusion of offline paper resources which are used alongside adult support for children who need additional support. The results were not conclusive as improvements were identified with inclusion and exclusion of these additional resources; future studies would need to explore the practicalities and benefits of their inclusion.

The review also provides evidence that Lexia could be an effective intervention for children with low socio-economic status as just over three-quarters of the participants were receiving free school meals which is a promising result considering the known trajectory of these students, and the increased cuts in school funding.

*Limitations*

Sources of bias may have impacted on the results; Lexia Learning is committed to producing products that are evidence-based and as a result funded five out of the six studies which were carried out by researchers employed by Lexia Learning. This is acknowledged in the studies but has been identified as a source of conflict within the pharmaceutical industry (Lundh, Lexchin, Mintzes, Schroll, & Bero, 2017), with further independent studies needed to corroborate the current findings. A further limitation is the small number of studies which were included in the review, and that the evidence for the effectiveness of the intervention in UK schools is limited to one study. The context, curriculum and school populations differ between the US and the UK, and as such the US studies are harder to generalise (Nind et al., 2005). The Education Endowment Fund (EEF) are currently investigating the effectiveness of Lexia through randomised control trials in more than 50 UK schools which will add valuable evidence to the effectiveness of the intervention in the UK.

Many of the studies included children who had English as an Additional Language, with promising results indicating that the intervention could support them (Kazakoff et al., 2017; Prescott et al., 2018. Schechter et al., 2015). However, the sample sizes of these groups were too small, or their results could not be extrapolated from the main population. Specific focus on this population was not included within the scope of this review but would be interesting to focus on in a future review. The limited impact of Lexia on children with poor phonological working memory (O'Callaghan et al., 2016) and the differentiated effects found by Baron et al., (2019) also points to future research assessing the effectiveness of Lexia for different groups of learners to ensure that schools make informed decisions about appropriate provision.

A further limitation of the reviewed studies was omitted evaluation of children and teacher's views about the intervention which could potentially impact on motivation and implementation.

Schechter et al., (2015) include teacher interviews which provide useful evidence, but even that does not highlight how teachers embed the intervention in their existing routine which has been found to be a key factor in successful intervention outcome (Outhwaite, Gulliford and Pitchford, 2019).

*Final conclusion*

Lexia is a promising intervention which could have educational significance due to the potential learning benefits for children, including those from families with lower socio-economic status. However, it still needs to be recommended with caution due to the majority of the research having been carried out in the US, and the lack of randomised control trials which limits understanding about the effectiveness of the intervention on reading outcomes.

## References

- Baisley, A., (2019). Analysis of School Funding Announcements September 2019. UNISON, 2019. Retrieved from: <https://www.unison.org.uk/content/uploads/2019/10/Analysis-of-Government-funding-announcements-20191014b.pdf>
- Barker, C., Pistrang, N., & Elliott, R. (2016). Research methods in clinical psychology : An introduction for students and practitioners / Chris Barker and Nancy Pistrang, University College London and Robert Elliott, University of Strathclyde. (3rd ed.).
- Baron, L., Hogan, T., Schechter, R., Hook, P., & Brooke, E. (2019). Can educational technology effectively differentiate instruction for reader profiles? *Reading and Writing*, 32(9), 2327-2352.
- Cheung, A., & Slavin, R. (2013). Effects of Educational Technology Applications on Reading Outcomes for Struggling Readers: A Best-Evidence Synthesis. *Reading Research Quarterly*, 48(3), 277-299.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers
- Duff, F., Hulme, C., Grainger, K., Hardwick, S., Miles, J., & Snowling, M. (2014). Reading and language intervention for children at risk of dyslexia: A randomised controlled trial. *Journal of Child Psychology and Psychiatry*, 55(11), 1234-1243.
- Dunlap, W. P., Cortina, J. M., Vaslow, J. B., and Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychology Methods* 1, 170–177.
- Gersten, R., Fuchs, L., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. (2005). Quality Indicators for Group Experimental and Quasi-Experimental Research in Special Education. *Exceptional Children*, 71(2), 149-164.
- Gibbs, S., & Bodman, S. (2014). *Phonological assessment battery* ( 2nd ed.). London, UK: GL Assessment.
- Gilbert, L., L, Teravainen, A, Clark, C, and Shaw, S. (2018) Literacy and life expectancy: An evidence review exploring the link between literacy and life expectancy in England through health and socioeconomic factors. National Literacy Trust, London.
- Gough, P. B., & Tunmer, W. E. (1986). Decoding, Reading, and Reading Disability. *Remedial and Special Education*, 7(1), 6–10.
- Harden, A., & Gough, D. (2012). Quality and Relevance Appraisal. In D. Gough, S. Oliver, & J. Thomas (Eds.). *An Introduction to Systematic Reviews*, 153–178, London: Sage
- Higgins, S. et al. (2017) Sutton Trust - Education Endowment Foundation Early Years Toolkit: Communication and language approaches, London: EEF.
- Kazakoff, E., Macaruso, R., & Hook, P. (2018). Efficacy of a blended learning approach to elementary school reading instruction for students who are English Learners. *Educational Technology Research and Development*, 66(2), 429-449.
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, 863.

- Lundh, A., Lexchin, J., Mintzes, B., Schroll, J., & Bero, L. (2017). Industry sponsorship and research outcome. *Cochrane Database Of Systematic Reviews*, 2(2), MR000033.
- Macaruso, P. & Rodman, A. (2011) Efficacy of Computer-Assisted Instruction for the Development of Early Literacy Skills in Young Children, *Reading Psychology*, 32:2, 172-196
- Macaruso, P., & Walker, A. (2008). The Efficacy of Computer-Assisted Instruction for Advancing Literacy Skills in Kindergarten Children. *Reading Psychology*, 29(3), 266-287.
- Macaruso, P., Wilkes, S., Franzén, S., & Schechter, R. (2019). Three-Year Longitudinal Study: Impact of a Blended Learning Program-Lexia® Core5® Reading-on Reading Gains in Low-SES Kindergarteners. *Computers in the Schools*, 36(1), 2-18.
- Mackay, T. (2002). The Future of Educational Psychology. *Educational Psychology in Practice*, 18(3), 245-253.
- McMurray, S. (2013). 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.
- Melby-Lervåg, M., Lyster, S., & Hulme, C. (2012). Phonological Skills and Their Role in Learning to Read: A Meta-Analytic Review. *Psychological Bulletin*, 138(2), 322-352.
- Melby-Lervåg, M., & Lervåg, A. (2014). Effects of Educational Interventions Targeting Reading Comprehension and Underlying Components. *Child Development Perspectives*, 8(2), 96-100.
- Mercer, N. and Fisher, E. (1997) Scaffolding through talk. In R. Wegerif and P. Scrimshaw (Eds.), *Computers and Talk in the Primary Classroom*, (pp.196-210). Clevedon: Multilingual Matters.
- NCS Pearson, Inc. (2012). aimsweb Technical Manual. Bloomington, MN: NCS Pearson, Inc.
- Nind, M., and Wearmouth, J. (2005) *A systematic review of pedagogical approaches that can effectively include children with special educational needs in mainstream classrooms with a particular focus on peer group interactive approaches* London, UK. EPPI-Centre Research Evidence in Education Library
- O'Brien, B., Habib, M., & Onnis, L. (2019). Technology-Based Tools for English Literacy Intervention: Examining Intervention Grain Size and Individual Differences. *Frontiers in Psychology*, 10, 2625.
- O'Callaghan, P., Mclvor, A., McVeigh, C., & 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. *The British Journal of Educational Psychology*, 86(4), 546-558.
- Outhwaite, L, Gulliford, A, & Pitchford, N. (2019). A new methodological approach for evaluating the impact of educational intervention implementation on learning outcomes. *International Journal of Research and Method in Education* (2019) (In Press)
- Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Malden, Mass. Oxford: Blackwell.
- Picton, I. (2019). Teachers' use of technology to support literacy in 2018. National Literacy Trust. London

Prescott, J., Bundschuh, K., Kazakoff, E., & Macaruso, P. (2018). Elementary school-wide implementation of a blended learning program for reading intervention. *The Journal of Educational Research*, 111(4), 497-506.

Rose, J. (2006) *Independent Review of the Teaching of Early Reading*. Nottingham: DfES Publications.

Schechter, R., Macaruso, P., Kazakoff, E., & 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.

Slavin, R., Lake, C., Davis, S., & Madden, N. (2011). Effective Programs for Struggling Readers: A Best-Evidence Synthesis. *Review of Educational Research*, 6(1), 1-26.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Williams, K. T. (2001). *Group reading assessment and diagnostic evaluation*. Circle Pines, MN: American Guidance Service.

Wood, D., Bruner, J., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Child Psychiatry*, 17, 89–100

Wright Group. (n.d.). *Reading, the Digital Classroom and LEAD 21*. Retrieved from <https://www.frontiercsd.org/cms/lib/NY19000265/centricity/domain/20/research-base-reading-digital-classroom.pdf>

Appendices

APPENDIX A

Table 1

*Studies Excluded After Full Text Screening*

Study	Exclusion Criteria
McMurray, S. (2013). An evaluation of the use of Lexia Reading software with children in Year 3, Northern Ireland (6- to 7-year olds. <i>Journal of Research in Special Educational Needs</i> , 13(1), 15-25.	Intervention: Studies that do not use Lexia® Core5®
Regan, K., Berkeley, S., Hughes, M., & Kirby, S. (2014). Effects of Computer-Assisted Instruction for Struggling Elementary Readers With Disabilities. <i>The Journal of Special Education</i> , 48(2)	
Macaruso, P. & Rodman, A. (2011) Benefits of Computer-Assisted Instruction to Support Reading Acquisition in English Language Learners, <i>Bilingual Research Journal</i> , 34:3, 301-315	
Macaruso, P. & Rodman, A. (2011) Efficacy of Computer-Assisted Instruction for the Development of Early Literacy Skills in Young Children, <i>Reading Psychology</i> , 32:2, 172-196	
Macaruso, P., Hook, P., & McCabe, R. (2006). The efficacy of computer-based supplementary phonics programs for advancing reading skills in at-risk elementary students. <i>Journal of Research in Reading</i> , 29(2), 162-172.	
Macaruso, P., & Rodman, A. (2009). Benefits of computer-assisted instruction for struggling readers in middle school. <i>European Journal of Special Needs Education</i> , 24(1), 103-113.)	
Macaruso, P., & Walker, A. (2008). The Efficacy of Computer-Assisted Instruction for Advancing Literacy Skills in Kindergarten Children. <i>Reading Psychology</i> , 29(3), 266-287.	
What Works Clearinghouse (2009). Lexia Reading. What Works Clearinghouse Intervention Report. What Works Clearinghouse.	Study design: Study does not collect any pre- and post-measures of intervention impact; meta-analyses or systematic reviews

**APPENDIX B**

*Description of Studies Included in Review*

Author	Country	Sample	Design	Intervention	Measures
O’Callaghan et al. (2016)	UK	Year 1 and 2 children in Northern Ireland  N = 98  N = 49 (Intervention) N = 49 (Wait-list control)	RCT with a 2-month follow-up and a no treatment wait-list control group	8-week block of daily access to Lexia Reading Core5	Phonological Assessment Battery 2 <sup>nd</sup> Edition(PhAB-2)  Included sub-tests: Blending, Non-Word Reading, Phoneme Segmentation and Phonological Working Memory
Baron et al. (2019)	USA	N = 594  3 <sup>rd</sup> grade (age 8)	One group pretest-posttest design	6 month block of Lexia Core 5 implemented alongside traditional classroom instruction as part of blended learning. Focus on effects of Lexia for different reader profiles (Mixed Deficits, Poor Comprehenders, Poor Decoders, Typical Readers)	aimsweb (Pearson, 2012) including <i>Reading Curriculum-Based Measure (R-CBM)</i> – measures oral word reading fluency and <i>Reading Maze Test (Maze)</i> – measures reading comprehension.  Improvement in Lexia Core 5 used as an additional measure  Measured at two time points (Fall and Spring).
Schechter et al. (2015)	USA	N = 83  N = 45 (Intervention) N = 38 (Control)  1 <sup>st</sup> and 2 <sup>nd</sup> grade children	RCT with no follow-up	8 month intervention using Lexia Core 5 alongside traditional classroom instruction as part of blended learning.	<i>Group Reading Assessment and Diagnostic Evaluation (GRADE)</i> (Williams, 2001)  Measures Vocabulary (Word Reading and Word Meaning) and Comprehension (Sentence

Comprehension and  
Passage  
Comprehension)

Prescott et al. (2018)	USA	N= 641  Kindergarten N = 84 1 <sup>st</sup> grade N = 111 2 <sup>nd</sup> grade N = 109 3 <sup>rd</sup> grade N = 126 4 <sup>th</sup> grade N = 103 5 <sup>th</sup> grade N = 108	One group pretest-posttest design	Lexia Core 5 used for average of 28.56 weeks alongside traditional classroom instruction as part of blended learning.	<i>Group Reading Assessment and Diagnostic Evaluation</i> (GRADE) (Williams, 2001)
Macaruso et al. (2019)	USA	Tracked children from start of Kindergarten to the end of 2 <sup>nd</sup> grade  1 <sup>st</sup> year of study (K) N= 83 2 <sup>nd</sup> year of study (1) N= 68 3 <sup>rd</sup> year of study (2) N= 68	Longitudinal study	Lexia Core 5 alongside traditional classroom instruction as part of blended learning.	<i>Group Reading Assessment and Diagnostic Evaluation</i> (GRADE) (Williams, 2001)
Kazakoff et al. (2017)	USA	Y1 study sample  N= 884 K = 82 1 = 244 2 = 146 3 = 170 4 = 140 5 = 102  Year 2 sample N= 130 (from original sample from previous year)	Quasi-experimental	Lexia Core 5 alongside traditional classroom instruction as part of blended learning.	aimsweb (Pearson, 2012)  K – used aimsweb Letter Naming Fluency (LNF); 1 <sup>st</sup> grade used Nonsense Word Fluency (NWF) and 2 <sup>nd</sup> through 5 <sup>th</sup> grade used R-CBM (Reading Curriculum Based measurement)

**APPENDIX C**

**Weighting of Studies**

**WoE A: Methodological Quality**

To assess methodological quality, a generic non review-specific judgement was established by evaluating each study against a published protocol. For the purpose of this review, the Gersten et al., (2005) Group Experimental and Quasi Experimental coding protocol was used in an amended form; amendments to specific quality indicators were made in order that it could be used for a universal population, and to include studies that did not have a control group (see Table 1).

Table 1

*Amendments and Rationale*

Original quality indicator	Amended quality indicator	Rationale
Was sufficient information provided to determine/confirm the participants demonstrated the disability(ies) or difficulties presented?	Does the study provide enough information about the population of participants to which the results can be generalised?	The coding protocol could be used for universal populations.
Were appropriate procedures used to increase the likelihood that relevant characteristics of participants in the sample were comparable across conditions if a control group was used?	Were appropriate procedures used to increase the likelihood that relevant characteristics of participants in the sample were comparable across conditions if a control group was used? Choose N/A if not control group.	N/A given as an option as not all studies had a control group.
Was sufficient information given characterizing the interventionists or teachers provided? Did it indicate whether they were comparable across conditions?	Was sufficient information given characterizing the interventionists or teachers provided?	Not all studies had a control condition, but the review was still interested in whether sufficient information had been provided.
Did the study provide not only internal consistency reliability but also test-retest reliability and interrater reliability (when appropriate) for outcome measures? Were data collectors and/or scorers blind to study conditions and equally (un) familiar to examinees across study conditions?	Did the study assess fidelity of implementation through inter-rater reliability and inter-observer agreement?	None of the included studies provided information about internal consistency reliability and test-retest reliability for outcome measures, but the review still wanted to be able to differentiate between the reviews that did provide interrater reliability and inter-observer agreement.

Scores are collated for both essential and desirable quality indicators with each indicator scoring one point. Studies need to achieve more than 9 essential quality indicators to score a point of 1, and 4 or more desirable quality indicators to score a point of 2. An overall score can be calculated and equated to a quality rating for WoE A; 3 = High Quality, 2 = Acceptable Quality; < 2 = Poor Quality. The results are summarised in table 2.

Table 2

*Weight of Evidence A Scores and Quality Rating*

Studies	Overall score	Quality Rating
O' Callaghan et al. (2016)	2	Acceptable
Baron et al. (2019)	1	Low
Schechter et al. (2015)	3	High
Prescott et al. (2018)	2	Acceptable
Macaruso et al. (2019)	2	Acceptable
Kazakoff et al. (2018)	1	Low

An example of the coding protocol used can be found in Appendix D.

**WoE B: Methodological Relevance to the Review Question**

A review-specific judgement was made about the relevance of the study design to the review question. Petticrew and Roberts (2003) state that studies that include a control group are best placed to answer questions about effectiveness, and that randomised control trials (RCTs) will have the most internal validity. Criteria was therefore developed which gave a greater weighting to studies that contained elements inherent in an RCT.

*Weight of Evidence B Criteria and Ratings*

High (3 points)	<ol style="list-style-type: none"> <li>1. Control group (with intervention group focused on effectiveness of Lexia)</li> <li>2. Random assignment of participants to treatment or control</li> <li>3. Pre and post and follow up scores</li> <li>4. Measures used to test effectiveness are clearly described</li> </ol>
-----------------	--

Medium (2 points)	<ol style="list-style-type: none"> <li>1. Control group</li> <li>2. Random assignment of participants to treatment or control</li> <li>3. Pre and post scores but no follow up</li> <li>4. Measures used to test effectiveness are clearly described</li> </ol>
Low (1 point)	<ol style="list-style-type: none"> <li>1. No control group</li> <li>2. No random assignment</li> <li>3. Pre and post scores are reported</li> <li>4. Measures used to test effectiveness reported</li> </ol>

*Note.* Each study requires 3 out of 4 criteria to achieve rating at each level.

Overall WoE B scores and quality ratings for each study are below in table 3.

Table 3

*Weight of Evidence B Scores and Quality Rating*

Studies	WoE B	Quality Rating
O' Callaghan et al. (2016)	3	High
Baron et al. (2019)	1	Low
Schechter et al. (2015)	3	High
Prescott et al. (2018)	1	Low
Macaruso et al. (2019)	1	Low
Kazakoff et al. (2018)	1	Low

**WoE C: Topic Relevance to the Review Question**

Each study was evaluated for how relevant the focus of the study was for the current review question. Rationale for defining criteria was based on four key elements which were deemed to be central to the review question:

*Four key elements*

1. Fidelity of Implementation
2. Focus on Lexia Core 5
3. UK-based sample
4. Utility as a universal school intervention

Table 4

*Weight of Evidence C Criteria and Ratings*

Weighting	Criteria
High (3) – Rating achieved if 6/8 criteria met	1. The study is primarily focused on a universal student population
Medium (2) – Rating achieved if 4/8 criteria met	2. The study is primarily focused on establishing the effectiveness of Lexia Core 5
Low (1) – Rating achieved if 2/8 criteria met	3. The study uses a UK-based sample
	4. Interviews with teachers are included which highlight any barriers to implementation of the intervention
	5. Information is provided about the training given to school staff implementing the intervention
	6. There is clarity over who is administering the standardised measures
	7. Consistency of intervention is measured through inter-observer agreement
	8. The study used offline and online materials

*Note.* Double weighting was given to Criteria 2.

Table 5

*Weight of Evidence C Scores and Quality Rating*

Studies	WoE C	Quality Rating
O’ Callaghan et al. (2016)	3	High
Baron et al. (2019)	1	Low
Schechter et al. (2015)	2	Medium
Prescott et al. (2018)	2	Medium
Macaruso et al. (2019)	2	Medium
Kazakoff et al. (2018)	1	Low

**WoE D: Overall Appraisal of a Study’s Findings in Relation to the Review Question**

An overall Weight of Evidence D rating and judgement was produced by combining all the scores from WoE A, B and C. Each dimension was judged as having equal weight with the final scores averaged to produce a final score that was translated into an overall judgement rating based on a score range: Low (1-1.6); Medium (1.7-2.3); High (2.4-3).

Table 6

*Weight of Evidence D Scores and Quality Rating*

Studies	WoE D	Quality Rating
O' Callaghan et al. (2016)	2.67	High
Baron et al. (2019)	1.0	Low
Schechter et al. (2015)	2.67	High
Prescott et al. (2018)	1.67	Low
Macaruso et al. (2019)	1.67	Low
Kazakoff et al. (2018)	1.0	Low

## APPENDIX D

### Example Coding Protocol

**Coding protocol:** Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C, & Innocenti, M. (2004). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children*, 71,149-164.

**Name of coder:** Natalie Jones

**Date:** 19.12.19

**Full Study Reference:** Baron, L., Hogan, T., Schechter, R., Hook, P., & Brooke, E. (2019). Can educational technology effectively differentiate instruction for reader profiles? *Reading and Writing*, 32(9), 2327-2352.

#### **Essential Quality Indicators - Quality indicators for describing participants**

Does the study provide enough information about the population of participants to which the results can be generalised?

Yes

No

N/A

Unknown/Unable to Code

Were appropriate procedures used to increase the likelihood that relevant characteristics of participants in the sample were comparable across conditions if a control group was used? Choose N/A if no control group.

Yes

No

N/A

Unknown/Unable to Code

Was sufficient information given characterizing the interventionists or teachers provided?

Yes

No

N/A

Unknown/Unable to Code

#### **Essential Quality Indicators - Quality indicators for Implementation of the Intervention and Description of Comparison Conditions**

Was the intervention clearly described and specified?

Yes

No

N/A

Unknown/Unable to Code

Was the fidelity of implementation described and assessed?

Yes

No

N/A

Unknown/Unable to Code

Was the nature of services provided in comparison conditions described?

Yes

No

N/A

Unknown/Unable to Code

***Essential Quality Indicators – Quality Indicators for Outcome Measures***

Were multiple measures used to provide an appropriate balance between measures closely aligned with the intervention and measures of generalised performance?

Yes

No

N/A

Unknown/Unable to Code

Were outcomes for capturing the intervention's effect measured at the appropriate times?

Yes

No

N/A

Unknown/Unable to Code

**Essential Quality Indicators – Quality Indicators for Data Analysis**

Were the data analysis techniques appropriately linked to key research questions and hypotheses? Were they appropriately linked to the limit of analysis in the study?

Yes

No

N/A

Unknown/Unable to Code

Did the research report include not only inferential statistics but also effect size calculations?

Yes

No

N/A

Unknown/Unable to Code

Essential Quality Indicators Total Score: 4

**Desirable Quality Indicators**

Was data available on attrition rates among intervention samples? Was severe overall attrition documented? If so, is attrition comparable across samples? Is overall attrition less than 30%?

Yes

No

N/A

Unknown/Unable to Code

Did the study assess fidelity of implementation through interrater reliability and inter-observer agreement?

Yes

No

N/A

Unknown/Unable to Code

Were outcomes for capturing the intervention's effect measured beyond an immediate post-test?

Yes

No

N/A

Unknown/Unable to Code

Was evidence of the criterion-related validity and construct validity of the measures provided?

Yes

No

N/A

Unknown/Unable to Code

Did the research team assess not only surface features of fidelity implementation (e.g. number of minutes allocated to the intervention or teacher/interventionist following procedures specified), but also examine quality of implementation?

Yes

No

N/A

Unknown/Unable to Code

Was any documentation of the nature of instruction or series provided in comparison conditions?

Yes

No

N/A

Unknown/Unable to Code

Did the research report include actual audio or videotape excerpts that capture the nature of the intervention?

Yes

No

N/A

Unknown/Unable to Code

Were results presented in a clear, coherent fashion?

Yes

No

N/A

Unknown/Unable to Code

	Total	Score
<b>Essential Quality Indicators</b> Total of >9 = Score 1 Total of <9 = Score 0	4	0
<b>Desirable Quality Indicators</b> Total of ≥4 = Score 2 Total of <4 = Score 1 Total 0 = Score 0	3	1
<b>Total Score</b> (3 = High Quality; 2 = Acceptable Quality; <2 = Poor Quality)		1

**Study Rating: Poor Quality**