Effective classroom instructions for primary literacy? A critical review of the causal evidence

Beng Huat See
Stephen Gorard

Durham University Evidence Centre for Education

Corresponding author:
Beng Huat See
Email: b.h.see@durham.ac.uk

Abstract

In the last two decades there has been a proliferation of proposals for instructional practices, all promising to improve pupils' academic attainment. Many of these approaches have not been robustly or independently evaluated. Schools, enthusiastic to improve their children's academic outcomes, may rely on some of the more popular but untested approaches. It is important to know which approaches are based on evidence of effectiveness, and which are not. Some approaches may be in widespread use and harming individuals' lives either directly, or at the expense of better approaches. This paper summarises and synthesises evidence from research worldwide, concerning literacy teaching in primary schooling, to identify robustly tested teaching pedagogies that were shown to be effective, especially for pupils struggling with reading and writing. The review only considers evidence that has the potential to demonstrate causation. Each included study is assessed in terms of its quality, and whether the intervention offered any benefit. The programmes identified as promising represent the most appropriate evidence-led ways of improving primary literacy, among those that had been evaluated at the time of writing. The strongest evidence of what works for children struggling with literacy includes a range of specific interventions such as Fresh Start, Butterfly phonics, Accelerated Reader, and Switch-on Reading or Reading Recovery.

It is important to know which approaches to teaching literacy are based on evidence, and which are not. Some may be in widespread use, harming individuals' progress even if only at the expense of better approaches. This paper synthesises evidence from research concerning literacy teaching in primary schooling to identify robustly tested teaching pedagogies shown to be effective, especially for pupils struggling with reading and writing. The review considers evidence that has the potential to demonstrate causation. Each study is assessed in terms of quality, and whether the intervention offered any benefit. The strongest evidence of what works for children struggling with literacy includes a range of specific interventions such as some phonics approaches, Accelerated Reader, and Switch-on Reading/Reading Recovery.

Introduction

Improving children's literacy is a focus of education policy in many countries, because literacy is an important life skill in its own right, and also essential for accessing the wider curriculum at school. It is therefore important to identify approaches to teaching literacy that work well, especially for those who struggle with reading and writing. This identification is not easy, and there are many unevidenced claims made by developers and advocates of particular approaches.

The last two decades have seen a proliferation of instructional practices available to schools, and promotion of the use of technology, software, and scripted curricula to aid teaching, all of which promise to improve pupils' literacy. Many have not been robustly or independently evaluated. Under pressure to produce results, schools may be misled by anything that promises success. It is important

that schools know which approaches are not evidence-informed, as these may actually do more harm than good, and which are more promising.

In England, the Education Endowment Foundation (EEF) Teaching and Learning Toolkit provides a potentially useful resource for schools on evidence-led approaches. The Toolkit summarises the evidence for some of the commonly known and tested approaches based on meta-analyses of prior research. Such meta-analyses average the 'effects' across all studies included, and these studies may vary considerably in terms of quality (e.g. level of attrition), phase of education and outcome measures (both type and quality). These quality factors can affect the apparent 'effect' size of an intervention. For example, larger studies are more likely to produce smaller effect sizes than smaller studies (Slavin and Smith 2009), and studies that use measures related to the intervention tend to show bigger effect sizes than those using treatment independent measures (Slavin and Madden 2011). Therefore, averaging effect sizes across studies can mask many issues relating to quality. The authors of the EEF Toolkit are aware of such difficulties, and are taking steps to address them (EEF 2018).

Evaluating single studies from scratch takes considerably longer than simply aggregating effect sizes. Consequently, the evidence for a number of widespread classroom practices remains unclear. This paper addresses this problem by considering the evidence from individual studies that evaluate common approaches used in the primary classroom for literacy (notably reading and writing skills), including those meta-analysed without quality control for the Toolkit, to provide a best evidence summary for teachers.

Improving primary literacy

Improving the literacy of children, especially those from disadvantaged backgrounds, has been a concern of successive governments in England. This concern is partly prompted by the relatively weak performance of children in international comparisons. Only 75% of children in England achieved the expected level in reading, and 78% in writing at the end of primary school (Department for Education [DfE] 2018). The figures are lower for poorer children, eligible for free school meals. This is a problem because literacy is such a fundamental gateway for further study. Pupils struggling to achieve their 'expected' reading level at primary school generally find it difficult to access the full secondary curriculum, which has implications for their subsequent learning (Wolf and Katzir-Cohn 2001, Pikulski and Chard 2005), and later life (Kuczera et al. 2016).

Previous reviews have identified a range of strategies for improving reading and writing. Wanzek and Vaughn (2007) and Wanzek et al. (2010, 2013) suggested that interventions for students with reading difficulties and disabilities should be provided early, and via small groups. Marulis and Neuman (2010) suggested that instruction involving trained adults in delivery was most effective in improving pre-K and kindergarten children's oral language development, and for children at risk of reading difficulties (Marulis and Neuman 2013). These studies used researcher-developed measures, with tests related to the intervention, so reducing the trustworthiness of the evidence.

Other reports have suggested that traditional methods of phonics and fluency are more effective than using cognitive approaches (Burns et al. 2016). Phonics is the teaching of letter and sound correspondences in an organised, regular, explicit and sequenced manner. It is a common approach to teaching literacy in primary schools in England, and one that is mandated by government. The Rose Report (Rose 2006) suggested that there was no good evidence that the analytic approach was effective, and so proposed greater use of phonics. The DfE (2015) claimed that there was substantial evidence that systematic synthetic phonics approach is the most effective method to teach children to read.

Other studies suggest that the teaching of morphology is beneficial for improving literacy, especially for weaker readers (Bowers et. al. 2010; Carlisle 2010). Goodwin and Ahn (2013) reported an overall "effect" size (ES) of 0.32 for morphological instruction in literacy. In each of these studies larger effect sizes were reported for quasi-experimental than for experimental studies, and for researcher-designed measures than for standardised ones. Again, this reduces the trustworthiness of the results.

Graham et al. (2012) synthesised evidence on improving writing, and concluded that most approaches were effective – including peer assistance (ES 0.89), self-regulated strategy (ES 1.17) and strategy instruction (ES 1.02). These are very large effect sizes, of a scale seldom found in education. A subsequent review (Graham et al. 2015) confirmed these findings, and added a few more approaches, including frequent writing (0.24). They also found that formal spelling instruction led to better literacy outcomes, at all levels of prior attainment.

With studies claiming that so many approaches work, it is difficult for teachers to use evidence to help decide which strategy they should employ. The problem with most previous reviews is that they do not consider the quality of the individual studies, but simply average effects across studies – i.e. averaging effects from strong studies (e.g. large randomised control trials), which are more likely to show smaller effects, with weak studies (with poor comparators, or intervention-aligned tests) known to be more susceptible to much bigger reported effect sizes. Evidence from such reviews can therefore be misleading by giving the impression that some programmes are more effective than they actually are.

This paper re-visits and reassesses the evidence for some common and popular approaches by looking at the quality of each individual study, and using a multi-factor method for judging the trustworthiness of each research finding.

Methods

The review robustly evaluates relevant empirical research, looking for evidence of a *causal* effect from teaching strategies or classroom processes on primary attainment in general. The full methods are in (Gorard et al. 2016). This paper presents the up-to-date results for literacy only. The review was conducted in a number of steps.

The search

The first search was completed at the end of 2016, limited to studies that were reported or published in English from 2000 onwards. In a subsequent search we updated our list with new studies published up to mid-2019. We sought research reports on children aged four to 12 (pre-school to end of primary phase, or immediate transition to secondary school) in mainstream education. We considered evaluations of teaching pedagogies/classroom practices, or interventions to enhance teaching instructions, that assessed the impact on pupils' academic outcomes. The full review considered all subjects, but for this paper we selected only the subset of results relevant to literacy. Our fuller results for some outcomes can be seen in Gorard et al. (2016) and Gorard et al. (2017). Unlike previous reviews completed by others, we considered only studies that have the potential to establish a causal impact. For this reason, we included only experimental or quasi-experimental designs and large-scale longitudinal studies, or similar. We excluded pieces that were non-empirical, in-depth case studies, with self-reported or teacher-reported or no pupil outcomes, or based outside the classroom.

The studies in this review were identified using two main search engines, which included the following sociological, psychological and educational electronic databases:

Web of Science – includes *Science Citation Index; Social Sciences Citation Index; Arts & Humanities Citation Index; Conference Proceedings Science Citation Index; Conference Proceedings Social Science & Humanities Citation Index; Book Citation Index - science, Book Citation Index - Social Science and Humanities, Emerging Sources Citation Index*

Ebscohost – includes *American doctoral dissertations, BEI, Child development & adolescent studies, ebook collection, Educational abstracts, Educational admin abstracts, ERIC, Library, information science & technology abstracts, MathSciNet via Ebscohost, Medline, PsyARTICLES, PSYCINFO)*

We also searched Google and Google Scholar to capture, in particular, unpublished studies that may have been missed otherwise, and the EEF Teaching and Learning Toolkit for relevant and current reports, added literature known to us from our previous studies, and followed up on studies identified in previous systematic reviews.

The electronic database search was conducted using the following keywords and their synonyms (for all outcomes not just literacy):

"Pedagog*" or "teaching effectiveness" or "teacher effectiveness" or "classroom practi*" or "classroom strategy* or "teaching strategy*" or "teaching approach*" or "teaching style" or "effective instruction" or "teach* practi*" or "teacher knowledge" or "teach* skill*" or "whole class teaching"
And
"Primary" or 'elementary" or "middle school" or "Key Stage 1" or "Key Stage 2" or "K-12" or "grade*" or "infant school" or "junior school" or "mobile children" or "migrant*"
And
"School outcomes" or "learning outcomes" or "academic performance" or "standardi* tests" or "exam*" or "key stage" or "grades" or "assessments" or "attainment" or "grade retention" or "grade point average"
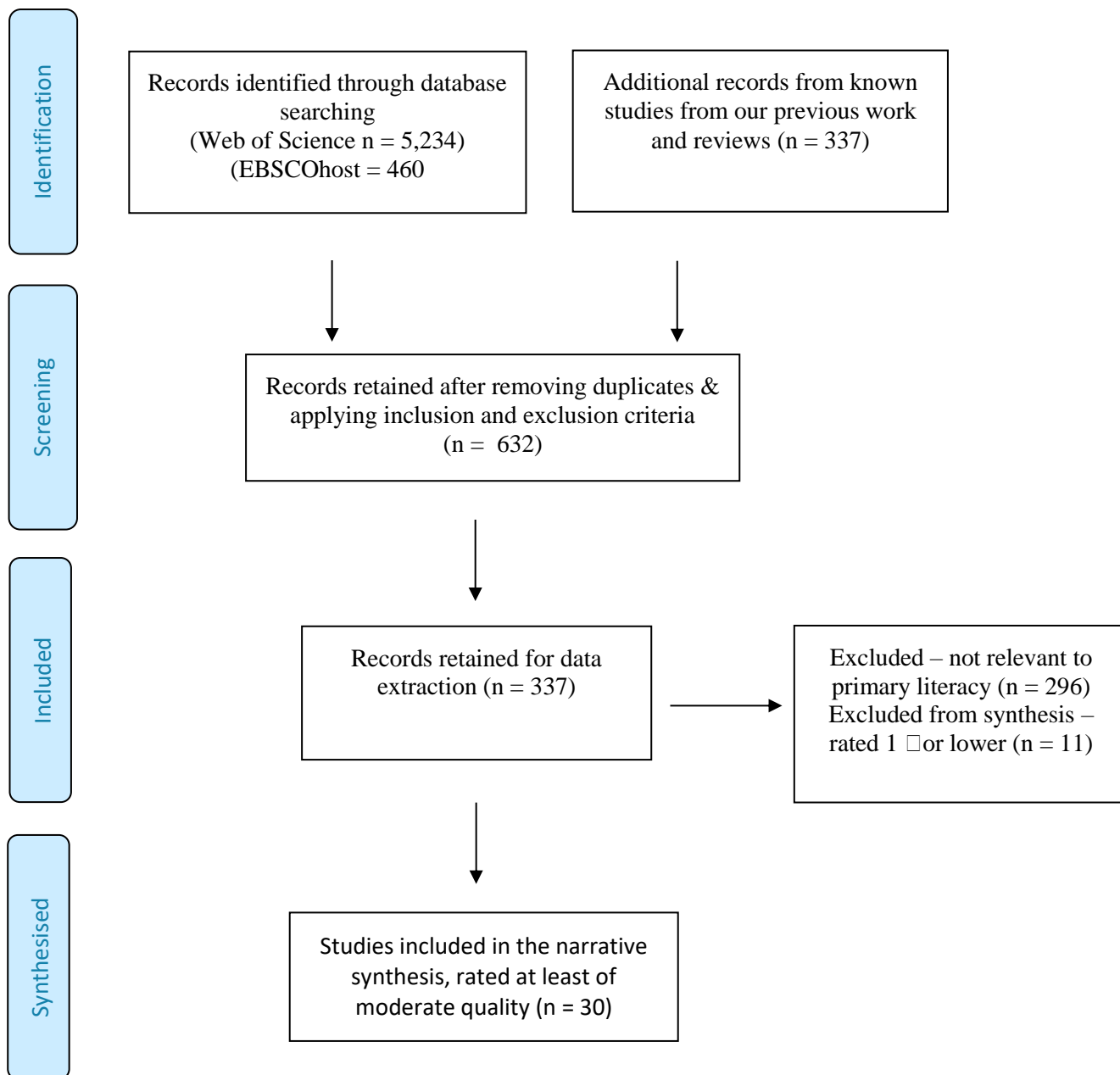And
"Trial" or "experiment" or "intervention" or "randomi* controlled trial" or "RCT" or "regression discontinuity" or "causal evidence"

The electronic searches together identified a total of 5,694 relevant studies.

Screening and data extraction

The retrieved studies were first screened for relevance by titles, and abstracts where available. After elimination of duplicates and applying the inclusion and exclusion criteria (above), a total of 632 were retained. Further reading of the full texts excluded another 295 for relevance, leaving 337 studies. We then extracted key information from each study, including the reported findings, research design, sample size and strategy, level of attrition, outcome measures, and other sources of bias. This information is necessary to make judgements about the trustworthiness of the evidence. In this paper we present only the stronger evidence, that is, those studies rated at least of moderate quality and relevant to primary literacy (Figure 1).

Prisma flow diagram for identification of studies

Records identified through database searching (Web of Science n = 5,234) (EBSCOhost = 460	Additional records from known studies from our previous work and reviews (n = 337)

Records retained after removing duplicates & applying inclusion and exclusion criteria (n = 632)

Records retained for data extraction (n = 337)

Excluded – not relevant to primary literacy (n = 296) Excluded from synthesis – rated 1 □or lower (n = 11)

Studies included in the narrative synthesis, rated at least of moderate quality (n = 30)

Identification

Screening

Included

Synthesised

Quality assessment

Our judgements of research quality, and of reliability between assessors, were aided by the 'sieve' developed by (Gorard et al. 2017). Each study was given a rating ranging from 0 (no weight can be placed on the study) to 4 🔒 (the most robust that could be expected in reality). This was based on five criteria: the design (a fair counterfactual), scale of study (size of smallest cell), bias through missing data, quality of data obtained (standardised tests versus developer constructed instruments) and other threats to validity (conflicts of interest). The judgement assumes that the study has reported these issues fully, clearly and without bias. If the study has not reported properly, and this impedes a fair judgement, it would automatically be rated as of lower quality. The ratings take no account of whether the intervention was deemed successful or not, or whether the report author claimed the intervention was effective.

To help judge the robustness of all non-zero findings and so whether missing data could alter the substantive findings, we also employ a sensitivity analysis based on the number of counterfactual cases needed to disturb the findings (NNTD). NNTD represents the number of cases with results antagonistic to the overall findings (by one standard deviation) that would be needed to make the overall findings zero. This is most easily assessed as the 'effect' size multiplied by the number of cases in the smallest group for any comparison. The bigger NNTD is, in relation to the number of missing values, the more secure any finding is (Gorard 2019). If NNTD is clearly larger than the number of missing values, then the overall result cannot be due solely to bias created by attrition.

Evidence synthesis

This was a very wide-ranging review, and collating the findings in meaningful ways was challenging. For this paper, we have selected only the findings that were judged moderately trustworthy (judged 2 🔒) or better, and only those concerned with improving literacy for primary school children. The small 1 🔒 studies tend to be biased by being disproportionately positive in outcomes (Gorard et al. 2017), and so added very little to our understanding. We then combined these better pieces into topics based on the relevant areas or approaches covered in the EEF Teaching and Learning website.

Presentation of results

Each topic is summarised in a table show how many studies are relevant, and the quality rating and outcome of each. This is followed by a table giving the reference, type of intervention, scale of the study, level of attrition, effect size, NNTD, and the overall quality rating, of each study.

Promising approaches

We begin our summary of the review results by looking at approaches that have been reasonably robustly evaluated and have shown some evidence of effectiveness for improving literacy, although they may not have been replicated yet. The first of these is phonics.

Phonics

Two 'experiments' had a considerable influence on the move to phonics after the Rose report (above). One involved 304 first year primary school children, allocated to three groups to receive different literacy interventions (Johnston and Watson 2004). But the groups were not randomly allocated, nor were they matched. Only the most disadvantaged pupils received the synthetic phonics intervention. This makes any findings insecure as the phonics group started from a lower base, so more improvement was possible in a short time. The second experiment by the same authors allocated the groups via matching, but involved only 92 first year pupils and these were divided into three groups (synthetic phonics, analytic phonics, and analytic phonics with phonological awareness training), leaving only

about 30 pupils per group. And the study was ended early for ethical reasons. Other commentators have suggested that the implementation of the three conditions may have led to bias (Wyse and Goswami 2008). Johnston et al. (2012) conducted two follow-up analyses using only some of their cases in an earlier evaluation in Scotland by comparing them with cases from England, unmatched on prior attainment. It is not clear why these weak studies have had the influence they had. Neither evaluation meets the minimum standard of evidence (2🔒) for this new review, and by the US What Works Clearinghouse [WWC] (2010).

There is actually little high quality research so far showing the impact (or not) of phonics teaching, compared to other plausible approaches. A review by Ehri et al. (2001) of effective reading programmes summarised the results of 68 experimental studies. It identified the key features of successful programmes as those that included teacher development and cooperative learning where children work with other children on structured activities and where there was a strong focus on phonics and phonics awareness, although a focus on phonics alone could not guarantee positive results. Thirteen studies for kindergarten children all reported strong positive outcomes. Most of these focus on phonics and cooperative learning and phonological awareness training. However, a combination of relatively few trials, poor evidence or poorly-reported methods, meant that the review result cannot be seen as definitive, especially in relation to exactly how phonics should be taught. A review of 20 RCTs on phonics interventions concluded that systematic phonics teaching was more effective than not using phonics, or using phonics non-systematically (Torgerson et al. 2006).

Most of the promising studies reported in two prior US reviews suggested that phonics training alone was not enough. The successful programmes included other elements such as cooperative learning or phonological awareness (Slavin et al. 2009, 2011). McArthur et al. (2012) found 12 studies using a variety of evaluation designs. Galuschka et al. (2014) reported 22 randomised controlled trials of phonics interventions. Both meta-analyses concluded that teaching phonics was more effective than other methods for reading accuracy, but not for spelling or reading fluency. However, few of these previous reviews have had a consistent approach to judging the quality of the underlying studies in order to calibrate their impact findings.

Despite all of these claims and counter-claims, our new subsequent review found only two studies of 3🔒 quality (Table 1). Both presented positive impacts, but neither assessed a classic *phonics-only* approach with primary pupils.

Table 1 - Quality and impact summary: studies of Phonics

	Effective	Ineffective/unclear
Higher quality (3 or 4🔒)	2	-
Medium quality (2🔒)	2	3

One was based on Fresh Start (FS), which is a 'systematic synthetic approach' produced by Read Write Inc., whose literacy programmes were cited by OFSTED (2010) as used by the 'best' performing schools. A recent RCT of Fresh Start suggested some promising results (Gorard et al. 2016). This was a school-led trial where 10 secondary schools came together to evaluate the approach for newly-arriving Year 7 pupils, with lower than expected levels of literacy. A total of 433 pupils were individually randomised to treatment and control conditions, in a wait-list design. There was slight initial imbalance as often occurs after randomisation, and this is why the study was not rated higher for quality. The intervention showed a small positive impact on reading comprehension (ES=0.24). A number of other studies of Fresh Start have reported positive outcomes (Brooks 2007), but these are generally very small (such as 29 cases split into two groups), with high dropout (Brooks et al. 2003 had 70% attrition from 500 initial cases), or used inappropriate designs such as before and after scores in one school (Lanes et al. 2005). Therefore, there is some promise but it needs replication or confirmation.

Another promising evaluation was based on online software called ABRACADABRA, developed in Canada for children with learning disabilities or those at risk of reading failure. ABRA is not a simple phonics intervention but has a toolkit of phonics, fluency, and comprehension skills. A large-scale efficacy trial of ABRA conducted in the UK used a 20-week intervention for 2,241 Year 1 pupils (age 5-6) involving 51 schools, with Progress in Reading Assessment (PIRA) to assess children's reading ability (McNally et al. 2018). The trial reported positive effects for reading using both the online (ES=0.14) and paper versions (ES=0.23). The results are made difficult to read because the key table showing the mean test scores did not include the standard deviations to allow for re-calculation of effect sizes. Attrition was 16%. There may have been a spillover effect, because control children in the treatment schools also performed better than those in control schools. Although the study is large, schools were first matched in pairs and then randomised in blocks, and pupils were divided into three conditions, all of which reduced the flexibility and so the scale of the randomisation. These are all reasons why the study quality was not rated higher.

We found three other evaluations of ABRA that were weaker in quality, because the trials were smaller, and used a diagnostic test called Group Reading Assessment and Diagnostic Evaluation (GRADE) for the outcome measure (Table 2). GRADE is closely aligned to the intervention, and so is not a fair test of literacy skills for the control group. A study by Savage et al. (2009) involved 144 pupils randomised to three groups - synthetic phonics, analytic phonics, and control. Both intervention groups used the on-line ABRA programme. Attrition was 18%, and higher for the weaker students (at baseline). The authors reported a positive impact from both intervention groups, but some results were actually negative – including reading fluency and comprehension in the synthetic group. The immediate effect size for the analytic group was 0.16.

A subsequent study by Savage et al. (2013) used 1,067 pupils, and 74 classes were randomised in pairs to either ABRA or control, and attrition was over 16%, because 12 classes did not provide post-test scores. Although the authors reported success, there were no benefits for Grades 1 and 2 pupils, and benefits in phonological blending, letter sound and phoneme segmentation fluency only for the reception year. A third ABRA study involved 360 primary pupils randomised in classes to treatment or control (Wolgemuth et al. 2013). The study reported positive outcomes, with an effect size of 0.36 but only when GRADE was the outcome measure. Using an independent measure of reading, Performance Indicators in Primary Schools (PIPS), the effect size was -0.21. This shows the importance of avoiding outcome measures linked closely to the intervention. Attrition was around 19%. Coupled with several smaller, weaker evaluations, again there is promise from ABRA which needs further work to confirm or refute.

Table 2 - Quality and impact detail: studies of Phonics

Reference	Intervention	Smallest cell	Attrition	ES	NNTD-attrition	Quality
Gorard et al. 2016	Fresh Start (Year 7)	216 pupils	3%	+0.24	45	3 🔒
McNally et al. 2018	ABRACADABRA	290 pupils	16%	+0.14 (on-line)	0	3 🔒
Savage et al. 2009	ABRACADABRA	44 pupils	18%	+0.16 (analytic only)	0	2 🔒
Savage et al. 2013	ABRACADABRA	37 classes	16%	0 (grades 1, 2)	0	2 🔒
Wolgemuth et al. 2013	ABRACADABRA	145 pupils	19%	-0.21 (PIPS) +0.36 (GRADE)	0	2 🔒

| King and Kasim 2015 | Rapid Phonics | 87 pupils | 13% | -0.05 | 0 | 2 🔒 |
| Merrell and Kasim 2015 | Butterfly Phonics | 155 pupils | 17% | 0.43 | 40 | 2 🔒 |

Two medium quality studies in England came to opposing conclusions about two slightly different phonics interventions. An evaluation of Rapid Phonics - a popular synthetic phonics programme used as a catch-up literacy intervention for pupils moving to secondary school – found no benefit. In fact the pupils in the treatment group did slightly worse than those in the control (King and Kasim 2015). However, Butterfly Phonics was found to be effective for pupils who were not achieving expected reading levels in the transition stage from primary to secondary school (Merrell and Kasim 2015).

There are, of course, a large number of smaller, weaker studies of phonics, with non-experimental designs, or that were conducted outside the classroom (Castiglioni-Spalten and Ehri 2003, Xue and Meisels 2004, Ryder et al. 2008, Kerins et al. 2010, Ritter et al. 2013). These may all add to our knowledge of some aspects of phonological awareness, but must be given little individual weight in a consideration of whether phonics teaching is likely to be effective in a standard primary classroom.

Overall, the best recent evidence on phonics is mixed but promising. Phonics works, at least for struggling readers, according to some measures, especially those most closely linked to phonics practice, and in some contexts. Some specific phonics products and protocols such as Fresh Start have stronger but un-replicated evidence (only Fresh Start and Butterfly phonics have individual evaluations with NNTD greater than zero).

Success for All

Success for all (SFA) is a whole school approach to literacy for primary school children. It is a multi-component training and support programme underpinned by co-operative learning strategies, and includes teaching synthetic phonics, ability grouping and ongoing assessment. The Allen report (2011) described SFA as one of the best early intervention programmes in the UK. A review, conducted by the programme developers, of studies mostly by the developers, reported positive results for struggling readers with a pooled effect size of around +0.5 (Slavin et al. 2011). Many of these studies used quasi-experimental or matched comparison designs, which tend to exaggerate the scale of the effects (Wilson and Lipsey 2007).

As with phonics, but to a lesser extent, there are several smaller, weaker studies (1 🔒 or worse) of SFA with positive or mixed results. Some had poor comparators, and suggested impact from phonics checks but not for other measures of literacy (e.g. Ross et al. 1998). In a study of eight schools using a matched comparison design, Tymms and Merrell (2001) found small positive effects for reading for some year groups only. We found no higher quality studies (rated 3 🔒 or 4 🔒). The best evidence comes from four medium quality studies, all suggesting some benefit from SFA (Table 3).

Table 3 - Quality and impact summary: studies of Success for All

	Effective	Ineffective/unknown
Higher quality (3 or 4 🔒)	-	-
Medium quality (2 🔒)	4	-

The first of these is a cluster RCT of 41 schools, of which six dropped out during the study, with 37% total attrition of pupils (Borman et al. 2007). Two cohorts of children within each school were randomised to receive the intervention or not, and followed for three years. Most of the children

dropping out were originally low attaining. After three years, positive outcomes were reported for word identification, comprehension, and especially phonics, but not reading efficiency (Table 4).

Quint et al. (2015) followed 2,956 Kindergarten children (age 5-6) from 37 schools up to 2nd Grade. Schools were randomly assigned to the SFA intervention (n=19) or control (n = 18). The findings suggested small gains for SFA. Overall attrition was 44%, and this coupled with the small number of cases to randomise reduced the strength of evidence to 2🔒.

An independent RCT conducted in England by Miller et al. (2017) shows no effect at the end of the first year of the trial (age 4-5), but a small positive effect (ES=+0.07) at the end of the second year (age 5-6), on the Woodcock reading mastery test, and a phonics check (but it must be recalled that phonics practice was part of the intervention). The trial was conducted in 54 schools and randomisation was at the school level. This study was rated 2🔒 because of the school-level randomisation (which meant fewer cases to randomise), relatively high attrition at 25% and the very low correlation (0.2) between the pre-test (BPVS) and the main post-test (Woodcock Reading Mastery Test).

A quasi-experimental study, conducted by the programme developers, involving 40 matched schools, reported positive effects for word identification and phonics (Tracey et al. 2014). Four schools dropped out, and there were then twice as many second language pupils in the SFA schools as in the control. This reduced the strength of evidence to 2🔒.

Table 4 - Quality and impact details: studies of Success for All

Reference	Intervention	Smallest cell	Attrition	ES	NNTD-attrition	Quality
Borman et al. 2007	SFA	17 schools	37%	0.22 Word Identification; 0.33 Phonics 0.21 Comprehension	0	2🔒
Quint et al. 2015	SFA	18 schools	44%	0.07 Word Identification 0.15 Phonics 0.07 Reading Efficiency 0.03 Comprehension	0	2🔒
Miller et al. 2017	SFA	27 schools	25%	0.07 Phonics	0	2🔒
Tracey et al. 2017	SFA	20 schools	25%	0.2 Word Identification 0.25 Phonics	0	2🔒

In summary, none of these studies had NNTD greater than zero, largely because of high attrition. In each study the strongest impact was reported for the phonics test, and phonics practice was part of the intervention. There is promise here for word identification, but larger, individually randomised trials and independently evaluated outcomes are needed.

Reading Recovery/Switch-On Reading

Reading Recovery (RR) is an intensive one-to-one intervention intended for the lowest performing 20% of first graders. It is widely used in the US, Australia, New Zealand and UK (Kelly et al. 2008). However, the evidence base does not really justify such widespread use. Tanner et al. (2011) compared 57 RR schools with 54 other schools, and reported that pupils at the RR schools performed better. However, the schools were not randomised to treatment, and baseline equivalence was not established. As with phonics and SFA, there have been many small or weak studies of RR, mostly yielding positive

outcomes (e.g. Holliman and Hurry 2013). The WWC (2013) found only four out of 78 evaluations of RR that met minimal evidence standards, and even these RCTs were small. Three suggested positive impacts. Our review, using a higher standard for causal evidence, found only one study of RR, of at least medium quality, and two further studies of a similar intervention called Switch-On Reading (Table 5).

Table 5 - Quality and impact summary: studies of Reading Recovery/Switch-On

	Effective	Ineffective/unknown
Higher quality (3 or 4 🔒)	1	-
Medium quality (2 🔒)	1	1

May et al. (2013) randomly assigned 866 low achieving first Grade students to small group RR or a control, and measured outcomes using the Iowa Test of Basic Skills (Table 6). Attrition was 17%, but the effect size reported was +0.68, which is quite substantial. In a follow-up, May et al. (2015) looked at one-to-one RR, and found an effect size of +0.47.

Table 6 - Quality and impact detail: studies of Reading Recovery/Switch-On

Reference	Intervention	Smallest cell	Attrition	ES	NNTD-attrition	Quality
May et al. 2013	Reading Recovery	433 pupils	17%	0.68	220	2 🔒
Gorard et al. 2015	Switch-On	157 pupils	1%	0.24	37	3 🔒
Patel, et al. 2017	Switch-on	341 pupils	10%	-0.04 Reading +0.05 Reading, writing	0	2 🔒

Switch-on Reading is derived at least partly from RR, but shorter in duration. There have been a few small, weaker research on Switch-on Reading (Coles 2012). Switch-on was evaluated in England at a larger scale by Gorard et al. (2015), with 157 pupils in each group, individually randomised to treatment or control. This study was based on pupils with poor reading skills as they made the transition from primary to secondary school. The outcomes were based on the independent New Group Reading Test pre- and post, attrition was only 1%, and there was an effect size of +0.24.

A follow-up effectiveness trial of Switch-on randomised 184 schools rather than pupils, and placed them in three groups rather than two (Patel et al. 2017) – meaning that the trial actually randomised fewer 'cases' to the smallest group than Gorard et al. (2015) did. There was a low correlation between the baseline measurement and the post-test, and higher attrition from the control group (11%) than the treatment groups. This study found little or no benefit from Switch-on.

Overall, treating these two related approaches as one, there is again considerable small-scale work, but the larger, better studies are generally positive. Two of these studies have NNTD clearly greater than zero.

Accelerated Reader

Accelerated Reader (AR) is a computerised reading programme which monitors and manages pupils' reading practices and encourages them in independent reading. In the UK, over 2,000 schools are using AR on a regular basis, which means that well over 400,000 students are reading what is recommended

in AR or what AR supports through quizzes (Topping 2014). However, it is not clear that the implementation of AR at such a large scale can be justified solely on the basis of the pre-existing evidence of effectiveness. There is a considerable research base on AR, making it one of the most researched interventions in which reading is practised through online resources. As usual, most of this research is small-scale, with high attrition, using AR-led measurements, or based on weak research designs for causal claims (Scott 1999, Pavonetti et al. 2003, Ysseldyke et al. 2003). Some of it also shows no benefit from using AR anyway, especially when using standardised tests that are not intervention-related (Mathis 1996, Facemire 2000, Duke 2011, Nichols 2013, Shannon et al. 2015).

There is also evidence of bias in some related research. Rudd and Wade's (2006) study, using matched comparison schools, showed that the average gains in reading from not using AR were greater than for the intervention schools, but this finding appears neither in their summary nor their conclusion. Instead, the authors reported that it needs "to be emphasized that there were improvements in average standardised test scores in the treatment schools for mathematics (both secondary and primary) and in the primary schools for reading" (p.51). The authors also did not report attrition clearly either at school or pupil level, and the reading attainment results are based on only 11 schools of the 21 originally allocated.

According to the WWC (2008) review, AR has no visible effect on reading fluency, a mixed effect on comprehension and a possible positive effect on reading achievement. Their results are based only two on studies that met WWC minimum standards, but these studies were still small and rather weak. Bullock (2005) used a total of 32 students. Brooks' (2007) meta-synthesis for AR found 47 studies conducted mostly in the US, but only two were selected for inclusion (Vollands et al. 1996, 1999). These studies reported positive effects. However, the cell sizes for comparison were only 11 pupils in one study, and 12 in the other. This is too small to be able to draw conclusions on the effectiveness of the intervention. It is not clear how the groups had been created, nor whether baseline equivalence was established between the treatment and control groups. And anyway in tests three months later, the control group had made more progress.

Our new review found only two studies of at least medium quality (Table 9).

Table 9 - Quality and impact summary: studies of Accelerated Reader

	Effective	Ineffective
Higher quality (3 or 4 🔒)	2	-
Medium quality (2 🔒)	-	-

In one study, 45 teachers (with 572 K-3 Grade students, aged 11-14, in 11 schools) were randomised to teach using AR or another commercially available reading programme (Ross et al. 2004). The results were analysed after one year, and there was 28% attrition. The authors reported a 'significant' impact on reading comprehension using the STAR reading test, with an effect size of +0.25, but WWC recalculated and reported that they found it was not statistically significant. Similarly, there was no significant effect on general reading achievement based on the STAR Early Literacy test for each year group. Also, the STAR tests are produced and marketed by Renaissance Learning as part of the AR programme itself (http://www.renlearn.co.uk/accelerated-reader/reports-and-data/).

Siddiqui et al. (2016) conducted an evaluation of AR in England involving 349 pupils in Year 7 who had not achieved the expected Level 4 in their Key Stage 2 results for English. They were individually randomised to treatment conditions. Attrition was 2%. The intervention group of 166 pupils outperformed the 183 control pupils on the independent New Group Reading Test (ES=+0.24).

Table 10 - Quality and impact detail: Accelerated Reader

Reference	Intervention	Smallest cell	Attrition	ES	NNTD-attrition	Quality
Ross et al. 2004	AR	286	28%	0.25	0	3 🔒
Siddiqui et al. 2016	AR	166	2%	0.24	78	3 🔒

Overall, the weight of evidence is that AR is a promising approach, and the two largest robust evaluations concur. However, it is not possible to identify which aspect of the programme is the driver(s) because AR is a multi-component intervention including explicit teaching, use of technology, differentiated instruction and self-regulated reading.

Approaches with little or no promise

It is important to understand what worked or might work, but it is also necessary to be aware of which interventions have no promise or might even be harmful. In a sense, all ineffective approaches are harmful because the opportunity and resources they use could have been used for something more likely to benefit that group of students and their life chances. The following are approaches with no solid evidence of effect.

Grammar for Writing

Grammar for Writing (GfW) is a literacy intervention intended to improve the writing skills of children by teaching writing that emphasises linguistic structures not grammatical rules. Other than work with secondary age pupils (Jones et al. 2013), we found two medium (or better) quality evaluations based on primary age pupils (Table 7).

One evaluation by Torgerson et al. (2014a) randomised two Year 6 classes (age 10-11) from 50 participating schools in England to either a whole class intervention or business-as-usual. Within the intervention class, eligible pupils were individually randomised to receive additional small group intervention or not, using minimisation to ensure that the groups were balanced. In other words, the groups were not really randomised, but partly matched on known characteristics. Overall attrition was around 22%. The primary outcome was extended writing using the Progress in English test, but 19% of pupils did not do the extended writing questions. This means that nearly half of those missing data were those who did not attempt the writing test, and for this reason the study is not rated higher than 2 🔒. The data shows little difference between intervention class and control (ES=+0.04). The effect was bigger for pupils receiving the additional small group intervention (ES=+0.20).

Table 7 - Quality and impact summary: studies on Grammar for Writing

	Effective	Ineffective/unknown
Higher quality (3 or 4 🔒)	-	-
Medium quality (2 🔒)	1	1

This initial work was scaled up for a trial involving Year 6 pupils in 155 schools (Tracey et al. 2019). The randomisation of schools was conducted in six batches, and differences between the groups in terms of observable characteristics were further minimised – again meaning that the cases were not really randomised, and the allocation by schools makes the process less flexible than the class randomisation used in Torgerson et al. (2014a). Attrition was 25%, and only schools that remained in the study were used in the analysis. The outcome measure was based on Key Stage 2 tests, rather than the Progress in

English, with a writing component, used by Torgerson et al. (2014a). The report does not include the standard deviations for the mean scores for each group, and so it is not possible for a reader to recalculate the effect sizes. The reported headline effect size is slightly negative.

Table 8 – Quality and impact detail: studies on Grammar for Writing

Reference	Intervention	Smallest cell	Attrition	ES	NNTD-attrition	Quality
Torgerson et al. 2014a	Grammar for Writing	408 pupils	22%	0.04 class 0.20 small group	0	2🔒
Tracey et al. 2019	Grammar for Writing	3,424 pupils	25%	-0.05	0	2🔒

In summary, relatively few robust evaluations have been done with Grammar for Writing in primary schools, and so the overall evidence is not promising.

READ 180

Read 180 is a small group reading programme designed for students not achieving the expected level of reading proficiency. It involves reading practice using a computer program, reading story books, and direct instruction in two 90-minute sessions. It has been evaluated in 156 studies listed by the WWC (2016). Many of these studies did not meet WWC standards for evidence (Interactive Inc. 2002), and those that did were often not able to provide clear evidence of benefit. Leaving aside the studies clearly based on older students (White et al. 2005), our review found the same three results (Table 11).

Table 11 - Quality and impact summary: Read180

	Effective	Ineffective/unclear
Higher quality (3 or 4🔒)	-	-
Medium quality (2🔒)	-	3

Fitzgerald and Hartry (2008) conducted an RCT with 484 children in four primary schools. Children were randomly assigned either to READ 180 as an afterschool programme or a regular afterschool programme. Attrition is not reported. The study found no effects on vocabulary, general literacy, reading fluency and spelling, and a small positive effect on reading comprehension but only for one cohort.

Kim et al. (2010) involved 264 struggling readers, and found no effects on any outcome measures, including reading fluency, comprehension and vocabulary, using norm-referenced and standardised tests. Attrition is not reported. The study involved randomly assigning poor readers from three elementary schools to READ180 or an alternative non-literacy focused after-school programme. Both groups scored below the proficiency level at the end of the intervention period.

In one of the largest studies of READ180, involving 5,551 students from Grade 6 to Grade 10 and independent standardised assessment, no differences were found overall, with a gain only in one middle school (Sprague et al. 2010). This study had a very high dropout rate of 55% which was not addressed by the complex analyses used by the authors.

Table 12 - Quality and impact details: Read180

Reference	Intervention	Smallest cell	Attrition	ES	NNTD-attrition	Quality

Kim et al. 2010	READ180	132 pupils	Unknown	0	0	2 🔒
Fitzgerald and Hartry 2008	READ 180	152 pupils	Unknown	0	0	2 🔒
Sprague et al. 2010	READ 180	2,775 pupils	55%	0	0	2 🔒

The overall evidence for READ180 is therefore mostly negative. It has had positive results mostly with assessments designed by the developers. It did not work in the few randomised control trials, or for those students most at risk. When positive results were found, they were for secondary school pupils or as an afterschool programme.

Project CRISS

Project CRISS (Creating Independence through Student-owned Strategies) is a programme where teachers model learning strategies for students to help develop independent learning. It is aimed at improving reading, writing and learning for 3rd to 12th Grade students. The strategies include monitoring learning, and building on prior knowledge with new information. The programme has been extensively research, but only two studies out of 31 met WWC minimal evidence standards (WWC 2010).

Table 13 - Quality and impact summary: Project CRISS

	Effective	Ineffective/unclear
Higher quality (3 or 4 🔒)	-	-
Medium quality (2 🔒)	1	1

An evaluation by Horsfall and Santa (1994) reported a positive effect on comprehension for students in Grades 4 and 6, based on an RCT involving 120 students in six intervention classrooms and 111 students in six control classrooms, after 18 weeks. The outcome was a teacher-developed 'free recall' comprehension test, which is not appropriate. There was dropout of around 20% from each class.

James-Burdumy et al. (2009) examined the impact of Project CRISS with 1,155 students attending 17 Project CRISS schools, and 1,183 students in comparator schools. The study found no benefit using a standardised norm-referenced diagnostic test (GRADE), or for reading comprehension.

Table 14 - Quality and impact details: Project CRISS

Reference	Intervention	Smallest cell	Attrition	ES	NNTD-attrition	Quality
Horsfall and Santa 1994	Project CRISS	111 pupils	20%	Mixed	-	2 🔒
James-Burdumy et al. 2009	Project CRISS	1,155 pupils	Unknown	0	0	2 🔒

Overall, the evidence of beneficial impact from Project CRISS is weak. Neither of the two best studies has a NNTD greater than zero.

IPEELL

IPEELL (Introduction, Point, Explain, Ending, Links and Language) is a writing intervention that uses pupils' memorable experiences together with self-regulation to develop their writing skills. IPEELL is related to Self-Regulated Strategy Development, which has also been studied, but with few robust evaluations so far (Sanders et al. 2019). We found two relevant IPEELL studies (Table 15).

Table 15 - Quality and impact summary: IPEELL

	Effective	Ineffective/unclear
Higher quality (3 or 4 🔒)	-	-
Medium quality (2 🔒)	1	1

The first was an efficacy trial (Torgerson et al. 2014b) involving children with poor writing in the transition phase (age 10 to 12), randomly assigned by school to intervention (12 primary schools) and control (11 schools). The headline outcome was pupils' writing performance measured on GL Assessment's Progress in English. Schools were allocated to treatment conditions matched on pupils' ethnicity and FSM-eligibility. Final analysis was based on 142 intervention and 119 control pupils (with at least 9% attrition). The study reported positive effects (ES +0.74) with larger benefits for free school meal (FSM) eligible children (ES +1.60). However, it is not possible to recalculate these because the report provided no mean scores or standard deviations.

Torgerson et al. (2018) conducted a larger effectiveness trial, involving 84 schools and 2,682 pupils, reporting after one year (with at least 13% attrition), and then after two years (24% attrition) once the pupils had moved to secondary schools. Here all pupils were included not just those with poor writing. After one year, IPEELL pupils made less progress (ES -0.09) based on KS2 writing scores, but after two years they made more progress (ES +0.11) using a past year KS2 writing assessment assessed by independent evaluators blind to group allocation. But after two years, the treatment group did worse in reading, perhaps because of the focus on writing.

Table 16 - Quality and impact details: IPEELL

Reference	Intervention	Smallest cell	Attrition	ES	NNTD-attrition	Quality
Torgerson et al. 2014	IPEELL	119 pupils	9%+	0.74	77	2 🔒
Torgerson et al. 2018	IPEELL	40 schools 38 schools	13% 24%	-0.09 one year +0.11 two years		2 🔒

Overall, there is not much evidence to support IPEELL, and the largest study has unclear results.

Writing Wings (WW)

Writing Wings is a structured writing programme in which children work in teams to help each other to write essays in various genres. We found two relevant studies (Table 17). Puma et al. (2007) evaluated 3,000 students in the 3rd to 5th Grades in 39 schools (Puma et al. 2007). No benefit was found for the writing ability of disadvantaged students (Table 18). Madden et al. (2011) conducted a randomised trial of Writing Wings which emphasised cooperative learning, with an added multimedia component. Thirty-two teachers from 22 schools were randomly assigned, with 922 children aged 8 to

10. The report has two different methods of analysis giving contradictory results - hierarchical linear modelling suggests no effect, and analysis of covariance shows small positive effect sizes for some outcomes. The overall results are therefore inconclusive.

Table 17 - Quality and impact summary: Writing Wings

	Effective	Ineffective/unclear
Higher quality (3 or 4 🔒)	-	-
Medium quality (2 🔒)	-	2

Table 18 - Quality and impact details: Studies of literacy

Reference	Intervention	Smallest cell	Attrition	ES	NNTD-attrition	Quality
Puma et al. 2007	Writing Wings	1,500 pupils	Unknown	0	0	2 🔒
Madden et al. 2011	Writing Wings	466 pupils	Unknown	0	0	2 🔒

There is no robust evidence that Writing Wings works.

One-off approaches

There is a number of approaches to teaching literacy in primary settings that are not based on named protocols or products, or which have only one study of direct relevance. As far as we can tell, these individual studies are not of any of the programmes covered so far, by another name. We found many studies, and most of these are too small or weak to address a causal question of effectiveness. Lee and Moore (2004) analysed the performance of struggling readers who were taught reading strategies, comparing one class of 27 pupils (9 of whom dropped out), with another class in the same school. Mason et al. (2013) had only 29 cases per group, with relatively high attrition. A study by Bjorn and Leppanen (2013) used 37 children from one school to evaluate software to improve reading fluency. Macedo-Rouet et al. (2013) based their small study on a matched comparison design with 12% attrition. Paris and Paris (2007) used an assessment that was only practised by the treatment group. Ignoring these 1 🔒 quality pieces, our review found three one-off studies (Table 19).

Table 19 - Quality and impact summary: One-off approaches

	Effective	Ineffective/unclear
Higher quality (3 or 4 🔒)	-	-
Medium quality (2 🔒)	3	-

6+1 Trait Writing is a supplemental writing programme that complements the schools' existing writing curricula. Coe et al. (2011) examined the impact for 4,161 Grade 5 students from 74 US schools (39 treatment, 35 control) using a cluster randomised experimental design. Schools were first matched by eligibility for free/reduced lunch, and one of each pair was then randomly assigned to treatment (2,230 pupils) or control (1,931). The outcome measure was Students' writing performance, rated on the six core characteristics of the 6+1 Trait Writing Model's definition of writing quality. So, the tests were similar to the intervention, making them unfair. Imputation was used to deal with missing data reported to be 6%. Treatment students increased their writing scores in a year - overall effect size (0.12 to 0.14).

Only three of the six outcome measures were reported as improved. Although this is recorded as "effective" in Table 19, this evidence is not sufficient alone to justify widespread use.

Mashburn et al. (2016) evaluated *Read It Again* – a curriculum which targets children's development of language and literacy - with 506 US pre-kindergarten pupils randomly assigned across three conditions: control, Read It Again, and RIA with expanded professional development. The relative advantage for RIA over control classrooms on each outcome was - print knowledge (0.07); alphabet knowledge (0.18); print concepts (0.25); definitional vocabulary (0.05); phonological awareness (0.02); and narrative language (0.15). Attrition is unknown, hence the 2 🔒 rating.

To evaluate the effectiveness of a *Three-Tier programme* used in *Reading Acceleration* (RAP), Houtveen and van de Grift (2012) conducted a quasi-experimental study, where 37 schools in the Netherlands already implementing the RAP programme were compared with schools randomly selected from the population. The sample included 1,021 children (567 treatment and 454 comparison group). A difference was found in the post-test for reading individual words and fluency (ES = 0.11 and ES = 0.46 respectively). The effects were sustained a year later. The comparison group had fewer struggling readers in the pre-test, so the results may be due to regression to the mean. The groups were unmatched and the experimental schools were already using the programme.

Table 20 - Quality and impact details: One-off approaches

Reference	Intervention	Smallest cell	Attrition	ES	NNTD-attrition	Quality
Coe et al. 2011	6+1 Trait Writing	1,931 pupils	6%	0.13	-	2 🔒
Mashburn et al. 2016	Read it Again	168 pupils	Unknown	Alphabet knowledge 0.18 Narrative language 0.15	-	2 🔒
Houtveen and van de Grift (2012)	Three-Tier	454 pupils	Unknown	Individual words 0.11 Fluency 0.46	-	2 🔒

In summary, none of the evidence for these three interventions is strong enough to be convincing in isolation. None of the studies has a NNTD minus attrition that is greater than zero.

Conclusions

Of course, like all reviews, this review cannot be assumed to include all relevant research. The issue is whether any of the research that was missed would change the substantive conclusions. This is the largest single-study review of this area, and so we are confident in the findings. However, it is important to note that we are not advocating any specific product or approach.

It is odd when conducting a review like this to note how few of the common approaches have been evaluated even with a medium quality trial. We found few (if any) evaluations of programmes that were replicated, or produced the same results when replicated. Similar programmes may be evaluated in multiple studies but for different populations, in different contexts, and using different outcome

measures, instruments, and duration. For this reason, we can only suggest programmes that have worked, and that we consider to be "best bets".

If improving literacy for disadvantaged pupils or struggling readers in the primary phase of education is the priority, then the best bets are those that have been robustly evaluated, shown a beneficial outcome for the treatment compared to a good counterfactual group, and for which the sensitivity test NNTD minus attrition yields a clearly positive result. This is a much safer way to envisage the most promising approaches than to rely on syntheses of meta-analyses aggregated without attention to quality. Very few of the approaches we looked at here, or in the wider review, met these stricter criteria, whatever their developers or sets of small, weak studies might portray.

The only approaches presented here that show promise as defined above are Fresh Start and Butterfly phonics (but no other phonics approaches as yet), Accelerated Reader which comes as a package including self-regulation teaching, and Switch-on Reading or Reading Recovery. It is not the purpose of this paper to endorse any individual programmes. The programmes identified as promising are those that we have found to be robustly evaluated, having passed stringent quality criteria, and so they represent the best evidence-led ways of improving primary literacy that we have at the time of writing.

It is important to stress that whatever programmes are used, for them to work schools need to stay faithful to the programme protocol. Some programmes may have failed to show a benefit because schools tried to modify the programme, either reducing the intensity or frequency of dosage, the structure of the programme (small groups instead of one-to-one) or applying it to different year groups. All of this can have an impact on the efficacy of any programme.

However, it must also be remembered that most plausible programmes do not work, and only around 10% to 15% of the most promising ideas for improving literacy have shown any benefits when tested rigorously (Gorard et al. 2017). This is shown again in our review. All of the rest are damaging, if only in the limited sense that their costs and the curriculum time spent on them inhibit the search for a better approach. Unless or until they are evaluated robustly and have results that stand up to sceptical scrutiny, all other approaches should be avoided.

References

Allen, G. (2011). *Early intervention: Smart investment, massive savings*. London: Cabinet Office.

Bjorn, P.M. and Leppanen, P.H. (2013). Accelerating decoding-related skills in poor readers learning a foreign language: a computer-based intervention. *Educational Psychology*, 33, 6, 671-689.

Borman, G.D., Slavin, R.E., Cheung, A.C.K., Chamberlain, A.M., Madden, N.A. and Chambers, B., 2007. Final reading outcomes of the national randomized field trial of Success for All. *American Education Research Journal,* 44, 3, 701–731.

Bowers, P.N., Kirby, J.R., Deacon, H. (2010). The effects of morphological instruction on literacy skills: A systematic review of literature. *Review of Educational Research*, 80, 2, 144-179.

Brooks, G. (2007) *What works for pupils with literacy difficulties? The effectiveness of intervention schemes*. London: DCSF Publications.

Brooks, G., Harman, J. and Harman, M. (2003) *Catching Up at Key Stage 3: an evaluation of the Ruth Miskin [RML2] pilot project 2002/2003*. A report to the Department for Education and Skills, Sheffield: University of Sheffield.

Bullock, J. (2005). *Effects of the Accelerated Reader on reading performance of third, fourth, and fifth-grade students in one western Oregon elementary school*. PhD Thesis, University of Oregon.

Burns, M., Peterson-Brown, S., Haegle, K., Rodrigues, M., Schmitt, B., Cooper, M., Clayton, K., Hutcheson, S., Conner, C., Hosp, J.L. and Vanderheyden, A.M. (2016). Meta-analysis of academic interventions derived from neuropsychological data. *School Psychology Quarterly*, 31,1, 28-42.

Carlisle, J. (2010). Effects of instruction in morphological awareness on literacy achievement: An integrative review. *Reading Research Quarterly*, 45, 4, 464-487.

Castiglioni-Spalten, M. and Ehri, L. (2003) Phonemic awareness instruction: Contribution of articulatory segmentation to novice beginners' reading and spelling. *Scientific Studies of Reading,* 7, 1, 25-52.

Cheung, A., Mak, B.S.Y., Abrami, P.C., Wade, A. & Lysenko, L. (2016). The effectiveness of the ABRACADABRA (ABRA) web-based literacy program on primary school students in Hong Kong. *Journal of Interactive Learning Research. 27,* 3, 219- 245

Coe, M., Hanita, M., Nishioka, V., Smiley, R. (2011) *An Investigation of the impact of the 6+1 Trait Writing Model on grade 5 student writing achievement: Final Report.* Washington DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Department of Education.

Coles, J. (2012) *An evaluation of the teaching assistant led Switch-on literacy intervention.* Unpublished MA Dissertation, University of London Institute of Education.

Comaskey, E. M., Savage, R. S. and Abrami, P. C. (2009). A randomised efficacy study of Web-based synthetic and analytic programmes among disadvantaged urban kindergarten children. *Journal of Research in Reading*, 32, 1, 92–108.

DfE (2015). *Reading: The next steps. Supporting higher standards in schools*. London: DfE.

DfE (2018) National curriculum assessments at key stage 2 in England, 2018, https://www.gov.uk/government/publications/national-curriculum-assessments-key-stage-2-2018-provisional/national-curriculum-assessments-at-key-stage-2-in-england-2018-provisional--2

Dianda, M., and Flaherty, J. (1995, April). *Effects of Success for All on the reading achievement of first graders in California bilingual programs*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Duke, J. (2011). *The Accelerated Reader Program in conjunction with best-practice reading instruction: The effects on elementary-school reading scores*. PhD. Thesis, Capella University, USA.

Education Endowment Foundation (2018) *Sutton Trust-Education Endowment Foundation Teaching and Learning Toolkit.* London: Education Endowment Foundation. https://educationendowmentfoundation.org.uk/resources/teaching-learning-toolkit

Ehri, L., Nunes, S., Stahl, S. and Willows, D. (2001). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. *Review of Educational Research*, 71, 3, 393-447.

Facemire, N. (2000) *The effect of the accelerated reader on the reading comprehension of third graders*, MA Dissertation, Salem-Teikyo University.

Fitzgerald, R. and Hartry, A. (2008). *What works in afterschool programs: The impact of a reading intervention on student achievement in the Brockton Public Schools (phase II).* Berkeley, CA: MPR Associates, Inc. and the National Partnership for Quality Afterschool Learning at SEDL.

Flis, A. (2006). *ABRACADABRA: The effectiveness of a computer-based reading intervention program for students at-risk of reading failure, and students learning English as a second language.* PhD Thesis, The University of British Columbia Open Collections.

Galuschka, K., Ise, E., Krick, K. and Schulte-Körne, G. (2014). Effectiveness of treatment approaches for children and adolescents with reading disabilities: a meta-analysis of randomized controlled trials. *PLoS One*, 26, 9, e105843.

Goodwin, A.P. and An, S. (2013). A meta-analysis of morphological interventions in English. *Education and Psychological Studies*, 17, 4, 257-285.

Gorard, S. and See, B.H. (2013). *Overcoming Disadvantage in Education*, London: Routledge

Gorard, S. (2014). A proposal for judging the trustworthiness of research findings. *Radical Statistics*, 110, 47–59

Gorard, S. (2019) Do we really need Confidence Intervals in the new statistics? *International Journal of Social Research Methodology*, 22, 3, 281-291.

Gorard, S., Siddiqui, N. and See, B.H. (2015). An evaluation of the 'Switch-on-Reading' literacy catch-up programme. *British Educational Research Journal*, 41, 4, 596-612.

Gorard, S., Siddiqui, N. and See, B.H. (2016). An evaluation of Fresh Start as a catch-up intervention: a trial conducted by teachers. *Educational Studies*, 42(1): 98-113

Gorard, S., See, B.H. and Siddiqui, N. (2017). *The trials of evidence-based education: The promises, opportunities and problems of trials in education*. London: Routledge

Graham, S., McKeown, D., Kiuhara, S., & Harris, K. R. (2012). Meta-analysis of writing instruction for students in elementary grades. *Journal of Educational Psychology,* 104, 4, 879-896.

Graham, S., Gillespie, A., and McKeown, D. (2013). Writing: Importance, development, and instruction. *Reading and Writing,* 26, 1, 1-15.

Holliman, A.J. and Hurry, J. (2013). The effects of Reading Recovery on children's literacy progress and special educational needs status: a three-year follow-up study. *Educational Psychology,* 33, 6, 719-733.

Horsfall, S. and Santa, C. (1994). *Project CRISS: Validation report for the program effectiveness panel*. Unpublished manuscript. In WWC (2010) *Project CRISS® (CReating Independence through Student-owned Strategies).* What Works Clearinghouse Intervention Report. Washington: US Department of Education, Institute of Education Sciences

Houtveen, T. and Van de Grift, W. (2012). Improving Reading Achievements of Struggling Learners. *School Effectiveness and School Improvement,* 23,1, 71-93.

Interactive Inc. (2002). *An efficacy study of READ 180: A print and electronic adaptive intervention program, grades 4 and above*. Ashland, VA. Author.

James-Burdumy, S., Mansfield, W., Deke, J., Carey, N., Lugo-Gil, J., Hershey, A., Douglas, A., Gersten, R., Newman-Gouchar, R., Dimino, J. and Faddis, B. (2009). *Effectiveness of selected supplemental reading comprehension interventions: impacts on a first cohort of fifth-grade students* (NCEE 2009-4032). Washington, DC: National Center for Educational Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education

Johnston, R.S., and Watson, J. (2004). Accelerating the development of reading, spelling and phonemic awareness. *Reading and Writing*, 17, 327–357.

Johnston, R.S., McGeown, S. and Watson, J.E. (2012). Long-term effects of synthetic versus analytic phonics teaching on the reading and spelling ability of 20-year old boys and girls. *Reading and Writing*, 25, 6, 1365-1384.

Jones, S., Myhill, D. and Bailey (2013) Grammar for Writing? An investigation into the effect of contextualised grammar teaching on student writing, *Reading and Writing Online*, 26, 1241–63

Kamps, D., Abbott, M., Greenwood, C., Wills, H., Veerkamp, M. and Kaufman, J. (2008). Effects of Small Group Reading Instruction and Curriculum Differences for Students Most at Risk in Kindergarten Two-Year Results for Secondary- and Tertiary-Level Interventions. *Journal of Learning Disabilities,* 41, 2, 101-114

Kerins, M., Trotter, D. and Schoenbrodt, L. (2010). Effects of a Tier 2 intervention on literacy measures: Lessons learned. *Child Language Teaching & Therapy*, 26, 3, 287-302.

Kim, J.S., Samson, J.F., Fitzgerald, R. and Hartry, A. (2010). A randomised experiment of a mixed-methods literacy intervention for struggling readers in grades 4-6: Effects on word reading efficiency, reading comprehension and vocabulary and oral reading fluency. *Reading and Writing: An interdisciplinary Journal,* 23, 1, 1109-1129.

King, B., and Kasim, A. (2015). *Evaluation of Rapid Phonics*, Education Endowment Foundation: London.

Kuczera, M., Field, S. and Windisch, H.C. (2016). *Building skills for all: A review of England*. Paris: OECD.

Lanes, D., Perkins, D., Whatmuff, T., Tarokh, H. and Vincent, R. (2005). *A survey of Leicester City Schools using the RML1 and RML2 literacy programme*. Leicester: Leicester City LEA (mimeograph).

Lee, E. and Moore, P. (2004). *Helping struggling readers comprehend information texts*. Paper presented at the International Association of School Librarianship conference, June 2004, Dublin, Ireland.

Macedo-Rouet, M., Braasch, J., Britt, M.A. and Rouet, J.F. (2013). Teaching Fourth and Fifth Graders to Evaluate Information Sources During Text Comprehension. *Cognition and Instruction*, 31, 2, 204-226.

Madden, N. A., Slavin, R. E., Karweit, N., Dolan, L., & Wasik, B. A. (1993). Success for All: Longitudinal effects of a restructuring program for inner-city elementary schools. *American Educational Research Journal*, 30, 1, 123–148.

Madden, N., Slavin, R., Logan. M. and Cheung, A. (2011). Effects of cooperative learning with embedded multimedia: A randomised experiment. *Effective Education*, 3, 1, 1-9.

Marulis, L. and Neuman, S. (2010). The Effects of Vocabulary Intervention on Young Children's Word Learning: A Meta-Analysis. *Review of Educational Research*, 80, 3, 300-335.

Mashburn, A., Justice, L.M., McGinty, A. and Slocum, L. (2016). The Impacts of a Scalable Intervention on the Language and Literacy Development of Rural Pre-Kindergartners. *Applied Developmental Science* 20, 1, 61-78.

Mason, L., Davison, M., Hammer, C., Miller, C. and Glutting, J. (2013). Knowledge, writing, and language outcomes for a reading comprehension and writing intervention. *Reading and Writing*, 26, 7, 1133-1158

Mathis, D. (1996). *The Effect of the Accelerated Reader Program on Reading Comprehension*, Report No. ED398555. ERIC Reproduction Services. US Department of Education, http://files.eric.ed.gov/fulltext/ED398555.pdf

May, H., Gray, A., Gillespie, J., Sirinides, P., Sam, C., Goldsworthy, H., Armijo, M. and Tognatta, N. (2013). *Evaluation of the i3 scale-up of Reading Recovery*, Consortium for Policy Research in Education, University of Delaware, Center for Research in Education and Social Policy (CRESP).

May, H., Gray, A., Sirinides, P., Goldsworthy, H., Armijo, M., Sam, C. Gillespie, J. and Tognatta, N. (2015). Year one results from the multisite randomized evaluation of the i3 scale-up of Reading Recovery. *American Educational Research Journal*, 52, 3, 547-581.

McNally, S., Ruiz-Valenzuela, J. and Rolfe, H. (2018). *ABRA: Online reading support. Evaluation report and executive summary*. London: Education Endowment Foundation.

Meisch, A., Hamilton, J., Chen, E. Quintanilla, P., Fong, P., Gray-Adams, K. and Thornton, N. (2011). *Striving readers study: Targeted and whole-school interventions – year 5*. Rockville, MD: Westat.

Merrell, C., and Kasim, A. (2015. *Evaluation of Butterfly Phonics.* London: Education Endowment Foundation: London.

Miller, S., Biggart, A., Sloan, S. and O'Hare, L. (2017). *Success for All. Evaluation report and executive summary*. London: Education Endowment Foundation.

Nichols, J. (2013). *Accelerated Reader and its effect on fifth-grade students' reading comprehension.* PhD Thesis, Liberty University.

Paris, A. H. and S. G. Paris (2007). Teaching narrative comprehension strategies to first graders. *Cognition and Instruction,* 25, 1, 1-44.

Patel, R., Jabin, N., Bussard, L., Cartagena, J., Haywood, S. and Lumpkin, M. (2017). *Switch-on effectiveness trial: Evaluation report and executive summary.* London: Education Endowment Foundation.

Pavonetti, L., Brimmer, K. and Cipielewski, J. (2000). *Accelerated Reader: What Are the Lasting Effects on the Reading Habits of Middle School Students Exposed to Accelerated Reader in Elementary Grades?* Paper presented at the Annual Meeting of the National Reading Conference 50th, Scottsdale, AZ, November 29-December 2, 2000)

Puma, M., Tarkow, A. and Puma, A. (2007). *The Challenge of Improving Children's Writing Ability: A Randomised Evaluation of Writing Wings.* Washington DC: Institute of Education Sciences, US Department of Education

Pikulski, J.J. and Chard, D.J. (2005). Fluency: Bridge between decoding and reading comprehension. *The Reading Teacher*, 58, 6, 510-519.

Quint, J., Zhu, P., Balu, R., Rappaport, S., and DeLaurentis, M. (2015). *Scaling up the Success for All model of school reform. Final report from the Investing in Innovation (i3) Evaluation*. New York: MDRC.

Ritter, M., Park, J., Saxon, T. and Colson, K. (2013). A Phonologically Based Intervention for School-Age Children With Language Impairment: Implications for Reading Achievement. *Journal of Literacy Research*, 45, 4, 356-385.

Rose, J. (2006). *Independent review of the teaching of early reading. Final report.* London: DfE.

Ross, S. M., & Casey, J. (1998). *Longitudinal study of student literacy achievement in different Title I school-wide programs in Fort Wayne community schools—Year 2: First grade results*. Memphis, TN: University of Memphis, Center for Research in Educational Policy.

Ross, S. M., Alberg, M., McNelis, M., & Rakow, J. (1997). *Evaluation of elementary school school-wide programs: Clover Park School District year 2: 1997–98.* Memphis: University of Memphis, Center for Research in Educational Policy.
American Educational Research Journal, 41(1): 191-229.
and phonemically based decoding skills as an intervention strategy for struggling readers
Collingwood, A., Chowdry, H., Greaves, E., Harrison, C., Johnson, G. and Purdon, S. (2011). *Evaluation of Every Child a Reader*. London: NatCen.
Evidence from the Early Childhood Longitudinal Study: Kindergarten Class of 1998-1999.
in whole language classrooms. *Reading and Writing*, 21,4, 349-369.
programmes for the elementary grades: A best evidence synthesis. *Review of Educational Research*, 79, 4, 1391-1466.

Ross, S. M., McNelis, M., Lewis, T., & Loomis, S. (1998). *Evaluation of Success for All programs: Little Rock School District year 1: 1997–1998*. Memphis, TN: The University of Memphis, Center for Research in Education Policy.

Ross, S., Nunnery, J. and Goldfeder, E. (2004). *A randomized experiment on the effects of Accelerated Reader/Reading Renaissance in an urban school district: Preliminary evaluation report.* Memphis, TN: The University of Memphis, Centre for Research in Educational Policy.

Rudd, P. and Wade, P. (2006). *Evaluation of Renaissance Learning Mathematics and Reading Programs in UK Specialist and Feeder Schools*, (Final Report). Slough: National Foundation for Educational Research.

Ryder, J., Tunmer, W. and Greaney, K. (2008). Explicit instruction in phonemic awareness

Sanders, S., Losinki, M., Ennis, R., White, W., Teagarden, J. and Lane, J. (2019) A meta-analysis of Self-Regulated Strategy Development Reading interventions to improve the reading comprehension of students with disabilities, *Reading and Writing Quarterly*, 35, 4, 339-353

Savage, R. S., Abrami, P., Hipps, G. and Deault, L. (2009). A randomized controlled trial study of the ABRACADABRA reading intervention programme in grade 1. *Journal of Educational Psychology*, 101, 3, 590-604.

Savage, R., Abrami, P.C., Piquette, N., Wood, E., Deleveaux, G., Sanghera-Sidhu, S. and Burgos, G. (2013). A (Pan-Canadian) cluster randomised control effectiveness trial of the ABRACADABRA web-based literacy programme. *Journal of Educational Psychology*, 105, 2, 310-328.

Scott, L. (1999). *The Accelerated Reader program, reading achievement, and attitudes of students with learning disabilities,* PhD Thesis, Georgia State University, Atlanta.

Shannon, L., Styers, M., Wilkerson, S., and Peery, E. (2015). Computer-assisted learning in elementary reading: A randomized control trial. *Computers in the Schools,* 32, 1, 20–34.

Siddiqui, N., Gorard, S. and See, BH (2016). Accelerated Reader as a literacy catch-up intervention during the primary to secondary school transition phase. *Educational Review,* 68, 2, 139-154.

Slavin, R. and Madden, N. (2011). *Understanding bias due to measures inherent to treatments in systematic reviews in education.* Paper presented at the annual meetings of the Society for Research on Effective Education, Crystal City, Virginia, March 3-4, 2008. https://files.eric.ed.gov/fulltext/ED527634.pdf

Slavin, R. and Smith, D. (2009). The Relationship between Sample Sizes and Effect Sizes in Systematic Reviews in Education. *Educational Evaluation and Policy Analysis*, 31, 4, 500-506. https://doi.org/10.3102/0162373709352369

Slavin, R. E., Lake, C., Davis, S., and Madden, N. A. (2011). Effective programs for struggling readers: A best-evidence synthesis. *Educational Research Review*, 6, 1–26.

Slavin, R., Lake, C., Chambers, B., Cheung, A. and Davis, S. (2009) Effective reading

Slavin, R., Wordsworth, J. and Jones-Hill, M. (2005). *Success for All in England: Implementation and outcomes of a comprehensive literacy reform for primary schools*. Baltimore, MD: Success for All Foundation.

Smith, L. J., Ross, S. M., Faulks, A., Casey, J., Shapiro, M., & Johnson, B. (1993). *1991–1992 Ft. Wayne, Indiana SFA results*. Memphis, TN: University of Memphis, Center for Research in Education Policy.

Sprague, K., Zaller, C., Kite, A., & Hussar, K. (2010). *Springfield-Chicopee School Districts Striving Readers (SR) program Year 3 report: Evaluation of implementation and impact.* Providence, RI: The Education Alliance at Brown University.

Tanner, E., Brown, A., Day, N., Kotecha, M., Low, N., Morrell, G., Turczuk, O., Brown, V.,

Topping, K. (2014). *What kids are reading: The book reading habits of students in British Schools 2014*: *An Independent Study*. Renaissance Learning Inc.: United Kingdom.

Torgerson, C., Ainsworth, H., Bell, K., Elliott, L., Gascoine, L., Hewitt, C., Kasim., Kokotsaki, D., and Torgerson, D. (2018). *Calderdale Excellence Partnership IPEEL: Evaluation report and executive summary.* London: Education Endowment Foundation.

Torgerson, C., Brooks, G. and Hall, J. (2006) *A Systematic Review of the Research Literature on the Use of Phonics in the Teaching of Reading and Spelling*. London: DfES, Research Report 711.

Torgerson, D., Torgerson, C., Ainsworth, H., Buckley. H., Heaps, C., Hewitt, C. and Mitchell, N. (2014b). *Improving writing quality: Evaluation report and executive summary*. London: Education Endowment Foundation.

Torgerson, D., Torgerson, C., Mitchell, N., Buckley, H., Ainsworth, H., Heaps, C. and Jefferson, L. (2014a). *Grammar for writing: Evaluation report and executive summary*. London: Education Endowment Foundation.

Tracey, L., Boehnke, J., Elliott, L.,Thorley, K., Ellison. S. and Bowyer-Crane, C. (2019). *Grammar for Writing: Evaluation report and executive summary*. London: Education Endowment Foundation.

Tracey, L., Chambers, B., Slavin, R. E., Hanley, P. and Cheung, A. (2014). *Success for All in England: Results from the Third Year of a National Evaluation*. *SAGE Open,* July-September, 1–10.

Tymms, P. and Merrell, C., (2001). *Success for All Quantitative Report: Second Year*. Durham: CEM, Durham University.

Vollands, S., Topping, K. and Evans, R. (1996*) Experimental evaluation of computer assisted self-assessment of reading comprehension: Effects on reading achievement and attitude*, ERIC Document, ED 408 567

Vollands, S., Topping, K. and Evans, R. (1999). Computerized self-assessment of reading comprehension with the Accelerated Reader: Action research. *Reading and Writing Quarterly*, 15, 3,197–211.

Wanzek, J., and Vaughn, S. (2007). Research-based implications from extensive early reading interventions. *School Psychology Review*, 36, 4, 541-561.

Wanzek, J., Vaughn, S., Lewis, N.S., Metz, C, Murray, C.S., Roberts, G. and Danielson, L, (2013). Extensive reading interventions for students with reading difficulties after Grade 3. *Review of Educational Research,* 83, 2, 1630195.

Wanzek, J., Wexler, J., Vaughn, S., Ciullo, S. (2010). Reading interventions for struggling readers in the upper elementary grades: A synthesis of 20 years of research. *Reading and Writing*. 23, 8, 889-912.

White, R., Williams, I, and Haslam, B.M. (2005). *Performance of District 23 students participating in Scholastic READ 180.* Washington, DC: Policy Studies Associates.

Wilson, S. and Lipsey, M., 2007. School-Based Interventions for Aggressive and Disruptive Behavior: Update of a Meta-Analysis. *American Journal of Preventive Medicine*, 33: 2S, 130–143.

Wolf, M., & Katzir-Cohen, T. (2001). Reading Fluency and Its Intervention. *Scientific Studies of Reading*, 5, 211-239.

Wolgemuth, J. R., Savage, R., Helmer, J., Lea, T., Harper, H., Chalkit, K., Bottrell, C., and Abrami, P. (2011). Using Computer-Based Instruction to Improve Indigenous Early Literacy in Northern Australia: A Quasi-Experimental Study. *Educational and Psychological Studies Faculty Publications,* 27, 4, 727-750.

Wolgemuth, J.r., Savage, R., Helmer, J., Harper, H., Lea, T., Abrai, P.C., Kirby, A., Chalkiti, K. Morris, P., Carapetis, J. and Louden, W., (2013). ABRACADABRA aids indigenous and non-indigenous early literacy in Australia: Evidence from a multisite randomised controlled trial. *Computers and Education,* 67, 250-264.

WWC (2008) *Accelerated Reader: WWC intervention report*, IES. http://ies.ed.gov/ncee/wwc/interventionreport.aspx?sid=12.

WWC (2010). *Project CRISS.* https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_projectcriss_061510.pdf.

WWC (2016). *Read 180 WWC intervention report*, IES. https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_read180_112916.pdf

Wyse, D. and Goswami, U. (2008). Synthetic phonics and the teaching of reading. *British Educational Research Journal*, 34, 6, 691-710.

Xue, Y. and Meisels, S. (2004) Early Literacy Instruction and Learning in Kindergarten:

Ysseldyke, J., Spicuzza, R., Kosciolek, S. and Boys, C. (2003). Effects of a learning information system on mathematics achievement and classroom structure. *Journal of Educational Research,* 96, 3, 163-173