



Cochrane
Library

Cochrane Database of Systematic Reviews

Phonics training for English-speaking poor readers (Review)

McArthur G, Sheehan Y, Badcock NA, Francis DA, Wang HC, Kohnen S, Banales E, Anandakumar T, Marinus E, Castles A

McArthur G, Sheehan Y, Badcock NA, Francis DA, Wang HC, Kohnen S, Banales E, Anandakumar T, Marinus E, Castles A.

Phonics training for English-speaking poor readers.

Cochrane Database of Systematic Reviews 2018, Issue 11. Art. No.: CD009115.

DOI: 10.1002/14651858.CD009115.pub3.

www.cochranelibrary.com

TABLE OF CONTENTS

HEADER	1
ABSTRACT	1
PLAIN LANGUAGE SUMMARY	2
SUMMARY OF FINDINGS FOR THE MAIN COMPARISON	4
BACKGROUND	7
OBJECTIVES	8
METHODS	8
Figure 1.	11
RESULTS	13
Figure 2.	14
Figure 3.	20
DISCUSSION	22
AUTHORS' CONCLUSIONS	25
ACKNOWLEDGEMENTS	25
REFERENCES	26
CHARACTERISTICS OF STUDIES	32
DATA AND ANALYSES	74
Analysis 1.1. Comparison 1 Phonics training versus control (random-effects model), Outcome 1 Mixed/regular word reading accuracy.	76
Analysis 1.2. Comparison 1 Phonics training versus control (random-effects model), Outcome 2 Non-word reading accuracy.	77
Analysis 1.3. Comparison 1 Phonics training versus control (random-effects model), Outcome 3 Irregular word reading accuracy.	78
Analysis 1.4. Comparison 1 Phonics training versus control (random-effects model), Outcome 4 Mixed/regular word reading fluency.	79
Analysis 1.5. Comparison 1 Phonics training versus control (random-effects model), Outcome 5 Non-word reading fluency.	80
Analysis 1.6. Comparison 1 Phonics training versus control (random-effects model), Outcome 6 Reading comprehension.	81
Analysis 1.7. Comparison 1 Phonics training versus control (random-effects model), Outcome 7 Spelling.	82
Analysis 1.8. Comparison 1 Phonics training versus control (random-effects model), Outcome 8 Letter-sound knowledge.	83
Analysis 1.9. Comparison 1 Phonics training versus control (random-effects model), Outcome 9 Phonological output.	84
Analysis 2.1. Comparison 2 Phonics training versus control: subgroup analyses (random-effects model), Outcome 1 Mixed/regular word reading accuracy.	85
Analysis 2.2. Comparison 2 Phonics training versus control: subgroup analyses (random-effects model), Outcome 2 Non-word reading accuracy.	88
Analysis 3.1. Comparison 3 Phonics training versus control: sensitivity analysis using the fixed-effect model, Outcome 1 Mixed/regular word reading accuracy.	90
Analysis 3.2. Comparison 3 Phonics training versus control: sensitivity analysis using the fixed-effect model, Outcome 2 Non-word reading accuracy.	91
Analysis 3.3. Comparison 3 Phonics training versus control: sensitivity analysis using the fixed-effect model, Outcome 3 Irregular word reading accuracy.	92
Analysis 3.4. Comparison 3 Phonics training versus control: sensitivity analysis using the fixed-effect model, Outcome 4 Mixed/regular word reading fluency.	93
Analysis 3.5. Comparison 3 Phonics training versus control: sensitivity analysis using the fixed-effect model, Outcome 5 Non-word reading fluency.	94
Analysis 3.6. Comparison 3 Phonics training versus control: sensitivity analysis using the fixed-effect model, Outcome 6 Reading comprehension.	95
Analysis 3.7. Comparison 3 Phonics training versus control: sensitivity analysis using the fixed-effect model, Outcome 7 Spelling.	96
Analysis 3.8. Comparison 3 Phonics training versus control: sensitivity analysis using the fixed-effect model, Outcome 8 Letter-sound knowledge.	97

Analysis 3.9. Comparison 3 Phonics training versus control: sensitivity analysis using the fixed-effect model, Outcome 9 Phonological output.	98
Analysis 4.1. Comparison 4 Phonics training versus control: sensitivity analysis with Hurford 1994 removed (random-effects model), Outcome 1 Mixed/regular word reading accuracy.	99
Analysis 4.2. Comparison 4 Phonics training versus control: sensitivity analysis with Hurford 1994 removed (random-effects model), Outcome 2 Non-word reading accuracy.	100
Analysis 5.1. Comparison 5 Phonics training versus control: sensitivity analysis with small studies removed (n < 11), Outcome 1 Mixed/regular word reading accuracy.	101
Analysis 5.2. Comparison 5 Phonics training versus control: sensitivity analysis with small studies removed (n < 11), Outcome 2 Non-word reading accuracy.	102
Analysis 5.3. Comparison 5 Phonics training versus control: sensitivity analysis with small studies removed (n < 11), Outcome 3 Irregular word reading accuracy.	103
Analysis 5.4. Comparison 5 Phonics training versus control: sensitivity analysis with small studies removed (n < 11), Outcome 4 Mixed/regular word reading fluency.	104
Analysis 5.5. Comparison 5 Phonics training versus control: sensitivity analysis with small studies removed (n < 11), Outcome 5 Non-word reading fluency.	105
Analysis 5.6. Comparison 5 Phonics training versus control: sensitivity analysis with small studies removed (n < 11), Outcome 6 Reading comprehension.	106
Analysis 5.7. Comparison 5 Phonics training versus control: sensitivity analysis with small studies removed (n < 11), Outcome 7 Spelling.	107
Analysis 5.8. Comparison 5 Phonics training versus control: sensitivity analysis with small studies removed (n < 11), Outcome 8 Letter-sound knowledge.	108
Analysis 5.9. Comparison 5 Phonics training versus control: sensitivity analysis with small studies removed (n < 11), Outcome 9 Phonological output.	109
ADDITIONAL TABLES	109
APPENDICES	125
WHAT'S NEW	135
HISTORY	135
CONTRIBUTIONS OF AUTHORS	136
DECLARATIONS OF INTEREST	136
SOURCES OF SUPPORT	137
DIFFERENCES BETWEEN PROTOCOL AND REVIEW	137
NOTES	138
INDEX TERMS	138

[Intervention Review]

Phonics training for English-speaking poor readers

Genevieve McArthur^{1,2}, Yumi Sheehan^{1,2}, Nicholas A Badcock^{1,2}, Deanna A Francis^{1,2}, Hua-Chen Wang^{1,2}, Saskia Kohnen^{1,2}, Erin Banales^{1,2}, Thushara Anandakumar^{1,2}, Eva Marinus^{1,2}, Anne Castles^{1,2}

¹Department of Cognitive Science, Macquarie University, Sydney, Australia. ²ARC Centre of Excellence in Cognition and its Disorders, Macquarie University, Sydney, Australia

Contact address: Genevieve McArthur, ARC Centre of Excellence in Cognition and its Disorders, Macquarie University, Sydney, Australia. genevieve.mcarthur@mq.edu.au.

Editorial group: Cochrane Developmental, Psychosocial and Learning Problems Group.

Publication status and date: New search for studies and content updated (no change to conclusions), published in Issue 11, 2018.

Citation: McArthur G, Sheehan Y, Badcock NA, Francis DA, Wang HC, Kohnen S, Banales E, Anandakumar T, Marinus E, Castles A. Phonics training for English-speaking poor readers. *Cochrane Database of Systematic Reviews* 2018, Issue 11. Art. No.: CD009115. DOI: 10.1002/14651858.CD009115.pub3.

Copyright © 2018 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

ABSTRACT

Background

The reading skills of 16% of children fall below the mean range for their age, and 5% of children have significant and severe reading problems. Phonics training is one of the most common reading treatments used with poor readers, particularly children.

Objectives

To measure the effect of phonics training and explore the impact of various factors, such as training duration and training group size, that might moderate the effect of phonics training on literacy-related skills in English-speaking poor readers.

Search methods

We searched CENTRAL, MEDLINE, Embase, 12 other databases, and three trials registers up to May 2018. We also searched reference lists of included studies and contacted experts in the field to identify additional studies.

Selection criteria

We included studies that used randomisation, quasi-randomisation, or minimisation to allocate participants to a phonics intervention group (phonics training only or phonics training plus one other literacy-related skill) or a control group (no training or non-literacy training). Participants were English-speaking poor readers with word reading one standard deviation below the appropriate level for their age (children, adolescents, and adults) or one grade or year below the appropriate level (children only), for no known reason. Participants had no known comorbid developmental disorder, or physical, neurological, or emotional problem.

Data collection and analysis

We used standard methodological procedures expected by Cochrane.

Main results

We included 14 studies with 923 participants in this review. Studies took place in Australia, Canada, the UK, and the USA. Six of the 14 included studies were funded by government agencies and one was funded by a university grant. The rest were funded by charitable foundations or trusts. Each study compared phonics training alone, or in conjunction with one other reading-related skill, to either no training (i.e. treatment as usual) or alternative training (e.g. maths). Participants were English-speaking children or adolescents, of low and middle socioeconomic status, whose reading was one year, one grade, or one standard deviation below the level expected for

their age or grade for no known reason. Phonics training varied between studies in intensity (up to four hours per week), duration (up to seven months), training group size (individual and small groups), and delivery (human and computer). We measured the effect of phonics training on seven primary outcomes (mixed/regular word reading accuracy, non-word reading accuracy, irregular word reading accuracy, mixed/regular word reading fluency, non-word reading fluency, reading comprehension, and spelling). We judged all studies to be at low risk of bias for most risk criteria, and used the GRADE approach to assess the quality of the evidence.

There was low-quality evidence that phonics training may have improved poor readers' accuracy for reading real and novel words that follow the letter-sound rules (standardised mean difference (SMD) 0.51, 95% confidence interval (CI) 0.13 to 0.90; 11 studies, 701 participants), and their accuracy for reading words that did not follow these rules (SMD 0.67, 95% CI 0.26 to 1.07; 10 studies, 682 participants). There was moderate-quality evidence that phonics training probably improved English-speaking poor readers' fluency for reading words that followed the letter-sounds rules (SMD 0.45, 95% CI 0.19 to 0.72; 4 studies, 224 participants), and non-word reading fluency (SMD 0.39, 95% CI 0.10 to 0.68; 3 studies, 188 participants), as well as their accuracy for reading words that did not follow these rules (SMD 0.84, 95% CI 0.30 to 1.39; 4 studies, 294 participants). In addition, there was low-quality evidence that phonics training may have improved poor readers' spelling (SMD 0.47, 95% CI -0.07 to 1.01; 3 studies, 158 participants), but only slightly improve their reading comprehension (SMD 0.28, 95% CI -0.07 to 0.62; 5 studies, 343 participants).

Authors' conclusions

Phonics training appears to be effective for improving literacy-related skills, particularly reading fluency of words and non-words, and accuracy of reading irregular words. More studies are needed to improve the precision of outcomes, including word and non-word reading accuracy, reading comprehension, spelling, letter-sound knowledge, and phonological output. More data are also needed to determine if phonics training in English-speaking poor readers is moderated by factors such as training type, intensity, duration, group size, or administrator.

PLAIN LANGUAGE SUMMARY

Phonics training for English-speaking poor readers

Review question

Does phonics training improve literacy-related skills in English-speaking poor readers.

Background

The reading skills of 16% of children fall below the average range for their age, and 5% of children have significant and severe reading problems. Poor reading is associated with higher risk of school dropout, as well as anxiety, depression, low self-concept, self-harm and suicide. Therefore, it is important to provide poor readers with early and effective help.

'Phonics' training is one of the most common reading treatments used with poor readers, particularly children. Phonics training teaches readers to: identify each letter or letter-cluster in a new word (e.g. S H I P); transpose each letter or letter-cluster into its corresponding speech sound ('sh' 'r' 'p'); and blend those speech sounds into a word ('ship').

Study characteristics

The search, updated in May 2018, identified 14 studies that tested phonics training in 923 English-speaking poor readers. The studies took place in Australia, Canada, the UK, and the USA. Six of the 14 included studies were funded by government agencies and one was funded by a university grant. The rest were funded by charitable foundations or trusts. Each study compared phonics training alone, or with one other reading-related skill, to either no training (i.e. treatment as usual) or alternative training (e.g. maths). Participants were English-speaking children or adolescents, of low and middle socioeconomic status, whose reading was one year, one grade, or one standard deviation (distance from the average) below the level expected for their age or grade for no known reason. Phonics training varied between studies in frequency (up to four hours per week), duration (up to seven months), training group size (individual and small groups), and delivery (human and computer). We measured the effect of phonics training on poor readers' ability to read words and novel words (non-words) accurately and fluently, as well as their comprehension of text, and their knowledge of letter-sound rules (letter-sound knowledge) and speech sounds (phonological output).

Key results

We found that phonics training in English-speaking poor readers probably improved irregular word reading accuracy, mixed/regular word reading fluency, and non-word reading fluency. It may also have improved mixed/regular word reading accuracy, non-word reading accuracy, reading comprehension, spelling, letter-sound knowledge, and phonological output.

Quality of the evidence

The overall quality of the evidence ranged from low to moderate. This was primarily due to large differences in the size of phonics-training effects between studies. More studies are needed to improve the precision of the outcomes.

Conclusions

The evidence suggests that phonics training can improve literacy in English-speaking poor readers. The positive effects of phonics training on all reading-related outcomes suggests that phonics training is not harmful for poor readers.

SUMMARY OF FINDINGS FOR THE MAIN COMPARISON *[Explanation]*

Phonics training versus control (no training or alternative training) for English-speaking poor readers						
Patient or population: English-speaking poor readers Setting: English-speaking countries Intervention: phonics Comparison: control (no training or alternative training)						
Outcomes	Illustrative comparative risks (SMD* 95% CI*)		Relative effect (95% CI)	N° of participants (studies)	Quality of the evidence (GRADE)*	Comments
	Assumed risk	Corresponding risk				
	Control (no training or alternative training)	Phonics training				
Mixed/regular word reading accuracy Assessed with: various scales Follow-up: immediate	-	The mean score in the intervention groups was 0.51 standard deviations higher (0.13 higher to 0.90 higher)	-	701 (11 studies)	⊕⊕○○ Low^a	A standard deviation of 0.51 represented a moderate effect between groups Phonics training “may improve” outcome (Ryan 2016).
Non-word reading accuracy Assessed with: various scales Follow-up: immediate	-	The mean score in the intervention groups was 0.67 standard deviations higher (0.26 higher to 1.07 higher)	-	682 (10 studies)	⊕⊕○○ Low^a	A standard deviation of 0.67 presented a moderate effect between groups Phonics training “may improve” outcome (Ryan 2016).
Irregular word reading accuracy Assessed with: various scales Follow-up: immediate	-	The mean score in the intervention groups was 0.84 standard deviations higher (0.30 higher to 1.39 higher)	-	294 (4 studies)	⊕⊕⊕○ Moderate^{a,c}	A standard deviation of 0.84 presented a large effect between groups Phonics training “prob-

					ably improves” outcome (Ryan 2016).	
Mixed/regular word reading fluency Assessed with: various scales Follow-up: immediate	-	The mean score in the intervention groups was 0.45 standard deviations higher (0.19 higher to 0.72 higher)	-	224 (4 studies)	⊕⊕⊕○ Moderate^b	A standard deviation of 0.45 presented a moderate effect between groups Phonics training “probably improves” outcome (Ryan 2016).
Non-word reading fluency Assessed with: various scales Follow-up: immediate	-	The mean score in the intervention groups was 0.39 standard deviations higher (0.10 higher to 0.68 higher)	-	188 (3 studies)	⊕⊕⊕○ Moderate^b	A standard deviation of 0.39 presented a moderate effect between groups Phonics training “probably improves” outcome (Ryan 2016).
Reading comprehension Assessed with: various scales Follow-up: immediate	-	The mean score in the intervention groups was 0.28 standard deviations higher (0.07 lower to 0.62 higher)	-	343 (5 studies)	⊕⊕○○ Low^a	A standard deviation of 0.28 presented a small effect between groups Phonics training “may improve” outcome (Ryan 2016).
Spelling Assessed with: various scales Follow-up: immediate	-	The mean score in the intervention groups was 0.47 standard deviations higher (0.07 lower to 1.01 higher)	-	158 (3 studies)	⊕⊕○○ Low^a	A standard deviation of 0.47 presented a moderate effect between groups Phonics training “may improve” outcome (Ryan 2016).

SMD: standardised mean difference. Different studies used different continuous measures. Thus, effect sizes are reflected by size of phonics training effect as indexed using SMDs. The results are expressed as standard deviation (SD) units. As a general rule, 0.2 SMD represents a small effect size, 0.5 a moderate effect size, and 0.8 a large effect size

CI: confidence interval.

GRADE: Working Group grades of evidence

High quality: we are very confident that the true effect lies close to the that of the estimate of the effect.

Moderate quality: we are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.

Low quality: our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of effect.

Very low quality: we have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect

^aDowngraded two levels due to very serious imprecision: very wide confidence intervals (greater than 0.6; [Schünemann 2011b](#)).

^bDowngraded one level due to serious imprecision: wide confidence intervals (0.3 to 0.6; [Schünemann 2011b](#)).

^cUpgraded one level due to large effect: SMD greater than 0.8 ([Ryan 2016](#)).

BACKGROUND

Description of the condition

The reading skills of 16% of children fall below the mean range for their age, and 5% of children have significant and severe reading problems (Shaywitz 1992). When children first learn to read, all written words are new to them. To read these words correctly, children need to learn how to: identify each letter in a word (e.g. S H I P); transpose each letter (I and P) or letter cluster (SH) into its correct speech sound using the letter-sound rules ('sh' 'i' 'p'); and blend these speech sounds into a word that can be said aloud ('ship'). These skills - which collectively can be termed 'phonics-based reading' - are detailed in a range of theoretical and computational models of reading (Coltheart 2001; Harm 1999; Perry 2007).

According to the "self-teaching hypothesis" (Share 1995), each time a new word is read via phonics-based reading, it forms and then strengthens a memory of that word's written form (e.g. SHIP). Subsequently, each time a person sees this word, this memory of the written form is activated, which, in turn, activates the meaning of that word (a boat), and the spoken version of that word ('ship'), which can be said aloud. Reading words via these processes is sometimes called 'sight-word reading'. These processes are also detailed in cognitive models of reading (Coltheart 2001; Harm 1999; Perry 2007).

Sight-word reading is particularly important for reading English for two reasons. First, it is faster and less effortful than reading via phonics-based reading skills (Ehri 2014; Weekes 1997). Second, a large proportion of written words in English contain letters that do not follow the letter-sound rules (i.e. they are 'irregular'; Vousden 2008); for example, we pronounce the ACH in YACHT like 'o' and not 'atch'. Most irregular words can be partially read with phonics-based reading since all irregular words have some letters that follow the letter-sound rules (e.g. Y and T in YACHT follow the letter-sound rules 'y' and 't'). However, to be read accurately, irregular words must be recognised individually via sight-word reading.

If a person has a problem with any of the processes involved in phonics-based reading or sight-word reading, then this will impair their ability to read. For example, if a person has poor phonics-based reading, they will have difficulty reading new words or names (e.g. EXPELLIARMUS) or non-words (i.e. nonsense words such as CHUB; Castles 1993) that follow the letter-sound rules. Alternatively, if a reader has poor sight-word reading, they should find it difficult to read irregular words accurately (such as YOU) and regular words efficiently (such as THINK; Castles 1993).

Poor reading is associated with higher risk of school dropout (Daniel 2006), as well as anxiety, depression, low self-concept, and self-harm and suicide (Alexander-Passe 2015; Carroll 2006; Maughan 2003; McArthur 2016).

Description of the intervention

This review focused on the most commonly investigated reading intervention for poor word readers: phonics. Phonics training teaches people to read via phonics-based reading, which depends upon the abilities to: identify each letter or letter-cluster in a word (e.g. S H I P); transpose each letter or letter cluster into its correct speech sound ('sh' 'i' 'p') using the letter-sound rules; and blend these speech sounds into a word that can be said aloud ('ship'; Savage 2018). Not all programmes that claim to be phonics programmes focus on phonics-based reading skills alone. Most programmes train numerous skills in combination with phonics, such as sight-word reading, phonological output, or reading comprehension. The results of these multi-faceted programmes are difficult to interpret because improvements in literacy-related outcomes could stem from phonics training, non-phonics training, or an interaction between the two. Therefore, the best way to test the efficacy of phonics training is to focus on 'pure' phonics programmes that train phonics-based reading skills alone.

How the intervention might work

According to evidence-based computational models of reading, phonics programmes should improve performance on tests of the individual processes that are involved in phonics-based reading (e.g. letter identification, letter-sound knowledge, sound blending), as well as on tests that tax all these processes simultaneously (such as regular word reading and non-word reading; Coltheart 2001; Harm 1999; Perry 2007). Since improvements in phonics-based reading should increase memories of whole written words, phonics should also improve performance on tests of processes involved in sight-word reading (e.g. memories of the written form of words, the meaning of words, and the spoken form of words) and on tests that tax these processes simultaneously (regular and irregular word reading). These gains in word reading may have knock-on effects on more complex literacy skills that depend on word reading such as reading comprehension and spelling.

The effect of phonics training on these reading skills may be influenced (i.e. moderated) by a number of factors. One factor is the type of training. As outlined above (Description of the intervention), most phonics interventions do not train phonics-based reading skills alone - 'pure' phonics interventions are rare. Thus, this review also considered phonics programmes that trained phonics-based reading skills plus one other literacy-related ability. The most common literacy-related skills that are trained alongside phonics reading skills are phoneme awareness (i.e. the ability to perceive, identify, discriminate, and manipulate speech sounds; see, for example, Blachman 2000; Hatcher 1994; Stahl 1994) and sight-word reading. We performed subgroup analyses to compare the effects of phonics training only, phonics training plus phoneme awareness training, and phonics training plus sight-word reading training on literacy outcomes.

A second factor that may moderate the effect of phonics training is training intensity. Previous studies conducted with typical readers have reported that phonics programmes that include a greater number of training sessions per week have a greater effect than programmes with fewer sessions (Bus 1999). Although logic would dictate that the same should be true for poor readers, this has yet to be tested empirically. We performed subgroup analyses that compared the efficacy of phonics programmes that involved up to two hours of training per week versus more than two hours of training per week.

A third moderating factor on phonics training may be the duration of the training period. We predicted that longer periods of phonics training would lead to greater reading gains than shorter programmes, and performed subgroup analyses to compare the efficacy of phonics programmes that were shorter than three months to those that were at least three months long.

A fourth factor that may moderate the effect of phonics is training group size. Previous research with typical readers has found that one-to-one phonics training is more effective than phonics training in a group (Ehri 2001). We expected the same to be true for poor readers and performed subgroup analyses to compare the effects of phonics in studies that conducted one-to-one training with poor readers and studies that trained small groups of poor readers.

A fifth moderating factor of phonics training may be the training administrator. One study reported that a reading training programme administered by a teacher is more effective than a programme administered by a computer (Dawson 2000), whereas another study found that delivering a reading programme via a computer alone is just as effective as delivering the same programme via a teacher and a computer (Torgesen 2010). In this review, we performed subgroup analyses to compare the effects of phonics training administered by a human versus phonics training administered via computer.

Why it is important to do this review

Many studies have tested the effect of phonics training in poor readers. Yet, surprisingly, there are few systematic reviews or meta-analyses on the effect of phonics training in people with poor reading. A very early review by Chall 1967 supported the use of phonics training for reading instruction, particularly for children from low socioeconomic backgrounds. However, this review did not measure the effect of phonics in poor readers specifically. The same was true for later meta-analyses by Elbaum 2000, Swanson 1999, and Therrien 2004. In contrast, three meta-analyses have measured the effect of phonics programmes specifically in poor readers (Ehri 2001; Galushka 2014; Suggate 2010). However, the review by Ehri was conducted well over a decade ago (Ehri 2001); the Galushka 2014 review focused on phonics interventions that simultaneously trained non-phonics skills; and the Suggate 2010 review excluded unpublished studies, focusing solely on children. Thus, in 2012, we conducted a review of the effects of specific phonics training

in poor readers regardless of age (McArthur 2012). The current review is an update of this work.

We are not aware of any studies that have tested the effect of phonics training on each of the skills involved in phonics-based reading and sight-word reading. It would be clinically and theoretically useful to look at the effects of phonics training on these specific processes (e.g. letter-sound knowledge, phonological output). It would also be informative to look at the efficacy of phonics training on reading skills that depend on these processes, such as regular-, irregular-, and non-word reading accuracy and fluency, as well as reading comprehension, and spelling.

Finally, we currently have little knowledge about the impact of moderating factors on phonics training in poor readers. For example, we do not know how intense or how long phonics training has to be, whether phonics training should be administered individually or in a small group, or if it should be delivered by a human or a computer. Again, this information will help teachers and therapists maximise the efficacy of their phonics training programmes.

OBJECTIVES

To measure the effect of phonics training and explore the impact of various factors, such as training duration and training group size, that might moderate the effect of phonics training on literacy-related skills in English-speaking poor readers.

METHODS

Criteria for considering studies for this review

Types of studies

Studies that allocated participants using random allocation, quasi-random allocation (e.g. defined by recruitment periods), or minimisation (i.e. minimised differences between groups for one or more factors).

Types of participants

Studies that recruited English-speaking children, adolescents, or adults, whose word reading was either one grade or one year (for children) or one standard deviation (SD) (for children, adolescents, and adults) below the appropriate level, for no known reason; that is, their poor reading did not stem from a comorbid developmental disorder (e.g. autism, language impairment, attention deficit hyperactivity disorder, attention deficit disorder); a physical problem (e.g. impaired vision); a neurological problem (e.g. brain

damage); or an emotional problem (e.g. long-term depression). This review did not exclude samples of poor word readers with a low intelligence quotient (IQ), since a discrepancy between IQ and reading is not predictive of prognosis or response to reading intervention (Fletcher 2005). Nor did it exclude participants based on age, gender, or socioeconomic status (SES), since response to reading intervention is not associated with a particular age, gender, or SES. This review was restricted to English-speaking poor readers because reading systems in different languages differ in the degree to which words can be read accurately using phonics-based reading skills. This review included studies that were conducted with poor readers who spoke English as their primary language at school or work, who lived in a country where English was the official language, and who were receiving phonics instruction in English. We excluded studies that included non-English speaking participants who had just arrived in an English-speaking country.

Types of interventions

Any phonics programme that trained a maximum of one other literacy-related skill (e.g. phoneme awareness training or sight-word training), compared with no treatment (effectively 'treatment as usual'), an alternate treatment (e.g. maths training), or a preintervention, double-baseline, no-training period.

Types of outcome measures

We measured the effect of phonics training on the primary and secondary outcomes listed below. The tests used by each study to measure the outcomes are summarised in Table 1.

Primary outcomes

1. Mixed/regular word reading accuracy (British Ability Scale - Word Reading).
2. Non-word reading accuracy (e.g. Castles and Coltheart 2 Test).
3. Irregular word reading accuracy (e.g. Castles and Coltheart 2 Test).
4. Mixed/regular word reading fluency (e.g. Test of Word Reading Efficiency - Sight Word Efficiency).
5. Non-word reading fluency (e.g. Test of Word Reading Efficiency - Phonemic Decoding Efficiency).
6. Reading comprehension (e.g. Neale Analysis of Reading Ability).
7. Spelling (e.g. Test of Written Spelling).

Secondary outcomes

1. Letter-sound knowledge (e.g. Woodcock Reading Mastery Test).
2. Phonological output (e.g. Children's Test of Phonological Processing).

Timing of outcome assessment

We measured the effect of phonics training in English-speaking poor readers immediately after training.

Search methods for identification of studies

We ran the searches for the original review in July 2012 (McArthur 2012), and for this update in February 2017 and May 2018. We used the Cochrane highly sensitive search strategy for identifying randomised trials in MEDLINE (Lefebvre 2011), and, where appropriate, adapted this strategy for use in the other databases. We limited the search by English language only when searching the ProQuest Dissertations and Theses database in May 2018. The search strategies for each database are reported in Appendix 1.

Electronic searches

We searched the following databases and trials registers.

1. Cochrane Central Register of Controlled Trials (CENTRAL; 2018, Issue 4), in the Cochrane Library, which includes the Cochrane Developmental, Psychosocial and Learning Problems Specialised Register (searched 9 May 2018).
2. MEDLINE Ovid (1946 to April week 4 2018).
3. MEDLINE In-Process & Other Non-Indexed Citations Ovid (searched 9 May 2018).
4. MEDLINE Epub Ahead of Print Ovid (searched 9 May 2018).
5. Embase Ovid (1980 to 2018 week 19).
6. *Cochrane Database of Systematic Reviews* (CDSR; 2018, Issue 4), part of the Cochrane Library (searched 9 May 2018).
7. Database of Abstracts of Reviews of Effects (DARE; 2015, Issue 2), part of the Cochrane Library (final issue of DARE, searched 15 February 2017).
8. ERIC EBSCOhost (Education Resources Information Center; 1966 to 9 May 2018).
9. PsycINFO Ovid (1966 to 9 May 2018).
10. CINAHL Plus EBSCOhost (Cumulative Index to Nursing and Allied Health Literature; 1966 to current).
11. Science Citation Index - EXPANDED Web of Science (SCI-EXPANDED; 1970 to 11 May 2018).
12. Social Science Citation Index Web of Science (SSCI; 1970 to 11 May 2018).
13. Conference Proceedings Citation Index - Science Web of Science (CPCI-S; 1990 to 11 May 2018).
14. Conference Proceedings Citation Index - Social Sciences & Humanities Web of Science (CPCI-SS&H; 1990 to 11 May 2018).
15. ZETOC (zetoc.jisc.ac.uk; searched 9 May 2018).
16. ClinicalTrials.gov (clinicaltrials.gov; searched 11 May 2018).

17. World Health Organization International Clinical Trials Registry Platform (WHO ICTRP; apps.who.int/trialsearch/; searched 11 May 2018).

18. metaRegister of Controlled Trials (www.isrctn.com; searched 11 May 2018).

19. ProQuest Dissertations and Theses Global (searched 24 May 2018).

Searching other resources

We examined the reference lists of published studies to identify further relevant studies. We contacted experts in the field and asked them to forward any published or unpublished, including ongoing, studies that we may have missed.

Data collection and analysis

Selection of studies

The team of review authors were divided into five pairs. Each pair of review authors was assigned one-fifth of the studies identified by the search terms. Each review author independently assessed their allocated studies against the inclusion criteria ([Criteria for considering studies for this review](#)). They then met with their corresponding partner to compare included and excluded studies, and discuss any disagreements. If no agreement was reached, the first author of this review made the final decision. Included studies undertaken by authors of this review were not assessed by said authors. Specifically, [McArthur 2015a](#) and [McArthur 2015b](#) were assessed by other review authors not engaged in these studies.

Data extraction and management

Five review author pairs (each author acting independently) extracted the data from each included study using the same data extraction form used for the original review. Data were collected on sample characteristics (including sample size); intervention characteristics (training type, training intensity, training duration, training group size, training administrator); and primary and secondary outcome measures (means, SDs, number of participants, and statistics). They then met with their corresponding partner to compare extracted data and resolve any disagreements. Any data missing from a study were dealt with using the procedures outlined in the [Dealing with missing data](#) section. The final data were entered into the [Data and analyses](#) section by the first author of this review (GMcA). It was double checked by the second author (YS) and then the fourth author (DF). Yet again, it is noteworthy that data extraction for included studies undertaken by authors of this review was not carried out by said authors (i.e. [McArthur 2015a](#); [McArthur 2015b](#)).

Assessment of risk of bias in included studies

Following the *Cochrane Handbook for Systematic Reviews of Interventions* ([Higgins 2011](#)), we rated each study at low, unclear, or high risk of bias, on the following seven domains: random sequence generation (selection bias), allocation concealment (selection bias), blinding of participants and personnel (performance bias), blinding of outcome assessment (detection bias), incomplete outcome data (attrition bias), selective reporting (reporting bias), and other bias. We presented our ratings in the 'Risk of bias' tables for each study (e.g. [Lovett 2000](#)), and graphically summarise them for all studies in [Figure 1](#).

Figure 1. Risk of bias summary: review authors' judgements about each risk of bias item for each included study.

	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding of participants and personnel (performance bias)	Blinding of outcome assessment (detection bias)	Incomplete outcome data (attrition bias)	Selective reporting (reporting bias)	Other bias
Barker 1995	+	+	?	?	?	?	+
Blythe 2006	+	+	+	?	+	?	+
Chen 2014	+	+	?	?	+	?	+
Ford 2009	+	+	?	+	+	?	+
Hurford 1994	+	+	?	+	+	?	+
Hurry 2007	+	+	?	+	+	?	+
Lewy 1997	+	+	+	?	+	?	+
Lewy 1999	+	+	+	?	+	?	+
Lovett 1990	+	+	?	+	+	?	+
Lovett 2000	+	+	?	+	+	?	+
McArthur 2015a	+	+	+	+	+	?	+
McArthur 2015b	+	+	+	+	+	?	+
Savage 2003	+	+	+	+	+	?	+
Savage 2005	+	+	+	+	+	?	+

Measures of treatment effect

Continuous data

All studies reported continuous data. Different studies used different tests to measure outcomes that used different scales (see [Table 1](#) for measures used in each study). Therefore, we used standardised mean differences (SMDs) with 95% confidence intervals (CIs) calculated from post-training group means and SDs for intervention and control groups. We considered SMDs of 0.20 to represent small, 0.50 to represent moderate, and 0.80 to represent large effects (Cohen 1988). In line with [Schünemann 2011a](#), we considered 95% CIs to be narrow if the range was around 0.10; medium if the range was around 0.30; and wide if over 0.60. These 95% CI ranges translate to high precision, moderate precision, and low precision in data. We considered intervention effects with a P value of 0.05 or less to be statistically reliable or statistically significant.

Unit of analysis issues

Multiple intervention groups

For the four studies that included more than one intervention group that received phonics training, we combined the post-training means, SDs, and numbers of participants (n) values of the groups ([Hurford 1994](#); [Levy 1997](#); [Levy 1999](#); [Savage 2003](#)). See [Characteristics of included studies](#) table for more details of these studies.

Three studies tested mixed/regular word reading or irregular word reading with two tests ([Barker 1995](#); [Lovett 1990](#); [Lovett 2000](#)). [Lovett 1990](#) tested word reading fluency with two tests (a mixed/regular word test and an irregular word test) and tested spelling with two tests (mixed/regular word spelling and irregular word spelling). For tests that used the same scale (e.g. Z scores that had a mean of 0 and SD of 1, or standard scores that had a mean of 100 and SD of 15), we calculated the average mean and average SD across the two tests. If the two tests used different scales (e.g. one test used Z scores and the other used standard scores), we:

1. calculated the SMDs for each test separately using the meta-analysis function in Review Manager 5 ([Review Manager 2014](#));
2. calculated the mean SMDs for the two tests;
3. removed the data entries for the two tests; and
4. inserted a new entry that used the mean SMD for the experimental group, 0 for the control mean, 1 for the SDs of both groups, and the n of the study.

In this update, we estimated effect sizes for mixed/regular and irregular words separately (see [Summary of findings for the main](#)

[comparison](#)). In the original review, we estimated the effect size for these outcomes combined, due to a lack of studies testing each outcome separately ([McArthur 2012](#)).

Dealing with missing data

If a study had missing data (e.g. means, SDs, amount of training, dropout rates), we requested that data from the corresponding author (see [Characteristics of included studies](#) table for details of communications). If this request failed, we contacted the coauthors. If a study excluded data for participants who failed to complete the training, or failed to adhere to the treatment programme, we asked the study authors for information about these cases. If an appeal for missing data did not result in a full data set, we only included data for participants whose results were known. We addressed the potential impact of any missing data in each study's 'Risk of bias' table ([Figure 1](#)) and the [Risk of bias in included studies](#) section.

Assessment of heterogeneity

We used a Chi² test with a P value of 0.10 to examine the degree of consistency in the effect sizes found by the included studies (i.e. heterogeneity; [Deeks 2011](#)). Further, we used the I² statistic (with a cut-off value of 70%) to estimate the percentage of variance in the effects owing to heterogeneity rather than chance.

Assessment of reporting biases

We used funnel plots to explore reporting bias for any outcome that had data from more than 10 studies which did not have similar standard errors for their effect sizes ([Sterne 2011](#)).

Data synthesis

We synthesised the outcomes of studies that used similar types of training (phonics alone, phonics plus one other literacy-related skill). We synthesised the outcomes of studies that trained children, adolescents, or adults, because there is no evidence that poor readers at different ages respond differently to different types of phonics training. See [Differences between protocol and review](#).

Subgroup analysis and investigation of heterogeneity

Subgroup analysis

The secondary aim of this review was to explore potential moderators on the efficacy of phonics training. We conducted subgroup analyses to test five potential moderators.

1. Training type (phonics alone, phonics and phoneme awareness, phonics and sight words).
2. Training intensity (less than two hours per week, at least two hours per week).
3. Training duration (less than three months, at least three months).
4. Training group size (one-to-one, small group).
5. Training administrator (human, computer).

Investigation of heterogeneity

We used a Chi² test with a P value of 0.10 to examine the degree of consistency in the effect sizes found by the included studies (i.e. heterogeneity; [Deeks 2011](#)). Further, we used the I² statistic (with a cut-off value of 70%) to estimate the percentage of variance in the effects owing to heterogeneity rather than chance. Where we found heterogeneity between studies (i.e. I² value greater than 70%), we: double-checked the data; reconsidered the validity and reliability of the measures; and examined outlier studies to see if there was an obvious reason for the outlying result and whether the outlying effects should be removed from each analysis.

To test the impact of heterogeneity on the outcomes, we calculated and compared (inverse variance) effect sizes using fixed-effect meta-analyses (which assumes the treatment effect is the same in each study) and random-effects meta-analyses (which assumes the treatment effect follows a distribution across studies; see [Table 2](#)). If the results for all outcomes proved similar, we reported the random-effects analyses since these adjust estimates to incorporate heterogeneity ([Deeks 2011](#)).

Sensitivity analysis

We conducted three sensitivity analyses:

1. removal of any studies with unclear random sequence generation;
2. removal of any studies with 10 or fewer participants in experimental and control groups ([Blythe 2006](#); [Chen 2014](#); [Ford 2009](#));
3. comparison of fixed-effect and random-effects meta-analyses for outcomes with high heterogeneity.

RESULTS

Description of studies

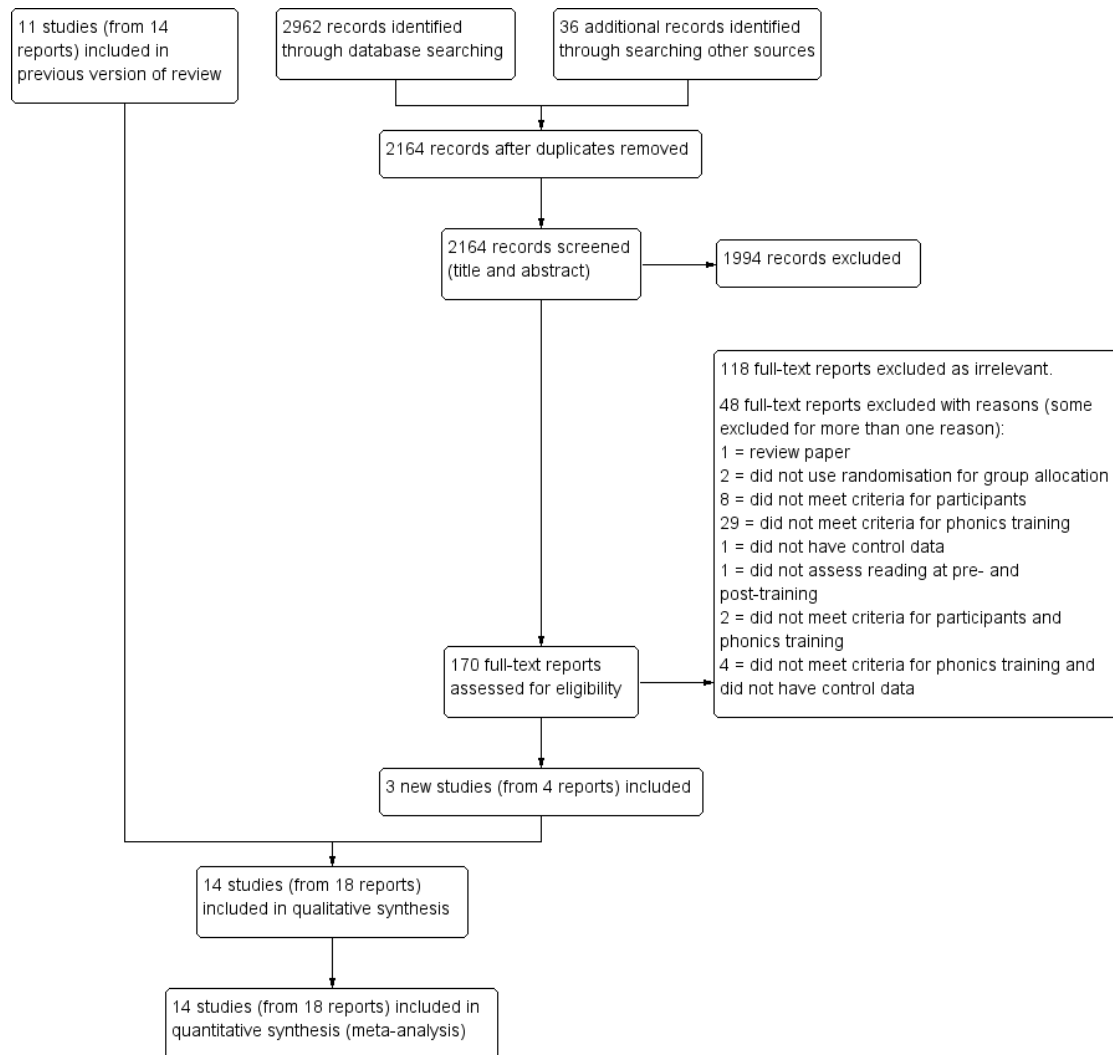
Results of the search

The searches for the original review, conducted in May 2011 and July 2012, resulted in 11 included studies (from 14 reports) ([McArthur 2012](#)).

For this update, our initial searches in February 2017 yielded a total of 2438 records. Having removed 830 duplicates, we screened the titles and abstracts of the remaining 1608 records against the inclusion criteria ([Criteria for considering studies for this review](#)), and identified 151 potentially relevant reports. Of these, we rejected 118 reports as irrelevant, formally excluded a further 29 with reasons (see [Excluded studies](#)), and included three new studies (from four reports) in the update. One of these reports, [Chen 2016](#), was a corrigendum for [Chen 2014](#), which supplied appendices that the journal had failed to publish with the 2014 paper. This corrigendum did not include additional data.

We ran top-up searches in May 2018 and identified 560 additional records. We removed four duplicates and screened the titles and abstracts of the 556 remaining records against our criteria ([Criteria for considering studies for this review](#)). We rejected 537 records as irrelevant, leaving 19 reports of potentially eligible studies. We formally excluded all 19 reports with reasons (see [Excluded studies](#)). Therefore, this review included 14 included studies (from 18 reports), three of which were new to this update. See [Figure 2](#).

Figure 2. Study flow diagram.



Included studies

Fourteen studies with 923 participants (see Table 3) met the inclusion criteria for this review (Barker 1995; Blythe 2006; Chen 2014; Ford 2009; Hurford 1994; Hurry 2007; Levy 1997; Levy 1999; Lovett 1990; Lovett 2000; McArthur 2015a; McArthur 2015b; Savage 2003; Savage 2005). Three of these studies (from four reports) were new to this update (Chen 2014; McArthur 2015a; McArthur 2015b). Three other papers described subsamples from Lovett 2000. Thus, this review included 14 studies from 18 reports.

Study design

All studies compared phonics training to a control group. All studies allocated participants using randomisation (Barker 1995; Blythe 2006; Chen 2014; Ford 2009; Hurford 1994; Hurry 2007; Levy 1997; Levy 1999; Lovett 1990; Lovett 2000; Savage 2003; Savage 2005), quasi-randomisation (McArthur 2015a), or minimisation (McArthur 2015b).

Location of studies

Five studies were carried out in Canada (Chen 2014; Levy 1997; Levy 1999; Lovett 1990; Lovett 2000), and three each in the UK (Hurry 2007; Savage 2003; Savage 2005), the USA (Barker 1995;

Ford 2009; Hurford 1994), and Australia (Blythe 2006; McArthur 2015a; McArthur 2015b).

Participants

See Table 3 for details about the participants in the individual studies. All studies reported details for participants who started the study rather than completed the study. However, it is noteworthy that all studies had very low or zero dropout rates, and dropout rates were similar across groups.

Reading ability

The criteria used to recruit poor readers differed between studies. Nine studies used some type of 'cut-off' point on a reading measure(s) such as: below the 40th, 20th, or 25th percentile (Barker 1995; Lovett 1990; Lovett 2000); a standard score less than 91 (Hurford 1994) or less than 90 (Levy 1999); less than seven words read correctly in an experimental measure (Levy 1997); a Z score or SD of -1 or less below the expected mean for age or grade (Chen 2014; McArthur 2015a; McArthur 2015b). Three studies recruited the poorest readers from a large sample of screened children (Hurry 2007; Savage 2003; Savage 2005), while two studies recruited children if they were participating in remedial reading at school (Blythe 2006; Ford 2009; note: data presented by these studies showed that the reading scores of these samples fell more than one SD below the level expected for their age, so the samples met the criteria for this review). Three studies also required participants to perform poorly on non-reading tests such as phoneme awareness tasks (Barker 1995; Savage 2003; Savage 2005). It is important to note that the different criteria used by each study did not determine its inclusion in this review, which used criteria broadly representative of those used by scientific reading researchers to identify studies with poor readers (see [Types of participants](#)).

Common exclusion criteria

Five of the 14 included studies reported criteria for exclusion from the study. The most common exclusion criteria were: low IQ scores (Blythe 2006; Lovett 1990); English as a second language (Lovett 1990; Lovett 2000); and a history of perceptual, psychological, or neurological problems (Lovett 1990; McArthur 2015a; McArthur 2015b). The remaining studies did not state exclusion criteria. Thus, differences between studies relating to exclusionary criteria added to the heterogeneity of samples both within and between studies.

Intelligence quotient

Two of the 14 included studies excluded participants with low IQ scores from their samples (Blythe 2006; Lovett 1990). Ten studies reported the verbal, non-verbal, or full IQ scores of their par-

ticipants (Barker 1995; Blythe 2006; Chen 2014; Hurford 1994; Hurry 2007; Levy 1999; Lovett 1990; Lovett 2000; McArthur 2015a; McArthur 2015b). The data suggested that most poor readers in these studies had IQ scores within or above the mean range.

English speakers (first or second language)

Four of the 14 included studies reported the ethnicity of their samples, which were either mixed (Ford 2009; Hurry 2007; Levy 1999), or predominantly white (Hurford 1994). One study reported participants being bilingual speakers of English and French (Chen 2014).

Age

Nine of the 14 included studies tested children aged between five and eight years (Barker 1995; Blythe 2006; Chen 2014; Hurford 1994; Hurry 2007; Levy 1997; Levy 1999; Savage 2003; Savage 2005). Four studies tested a slightly older and broader age group: seven to 13 years (Lovett 1990; Lovett 2000; McArthur 2015a; McArthur 2015b). One study tested adolescents (Ford 2009).

Gender

Eight of the 14 included studies tested about equal numbers of females and males (Ford 2009; Hurford 1994; Hurry 2007; Levy 1997; Levy 1999; McArthur 2015b; Savage 2003; Savage 2005). Four studies tested a larger proportion of males (around 63% to 75%) than females (around 25% to 36%) (Blythe 2006; Lovett 1990; Lovett 2000; McArthur 2015a). One study tested a larger proportion of females than males (Chen 2014). One study did not report the numbers of girls and boys in the study (Barker 1995).

Socioeconomic status

Three of the 14 included studies reported the SES of their sample, which was either low SES (Ford 2009; Savage 2005) or middle SES (Lovett 1990).

Interventions

Four of the 14 included studies included more than one phonics training group (Hurford 1994; Levy 1997; Levy 1999; Savage 2003); in these cases, we merged the data from the phonics training groups. Eight of the 14 studies included additional non-phonics training groups that were not included in the review (Barker 1995; Hurry 2007; Levy 1997; Levy 1999; Lovett 1990; Lovett 2000; McArthur 2015a; McArthur 2015b). See [Characteristics of included studies](#) tables for details.

Studies in this review used training programmes that differed in training: type (phonics only, phonics and phoneme awareness training, or phonics and sight-word training); intensity (less than

two hours per week or at least two hours per week); duration (less than three months or at least three months); group size (one-to-one or small group); and administrator (human or computer). These five categories corresponded to our five subgroup analyses ([Subgroup analysis and investigation of heterogeneity](#)). The studies that fall into each of the subgroups are summarised in [Table 4](#) and are discussed, in turn, below.

Training type

Phonics only

Five of the 14 included studies trained poor readers with a programme that focused on training children to read using phonics-based reading skills ([Barker 1995](#); [Levy 1997](#); [Levy 1999](#); [McArthur 2015a](#); [McArthur 2015b](#)). [Barker 1995](#) used the Hint and Hunt programme that taught children to read with the letter-sound rules for short vowel sounds. [Levy 1997](#) and [Levy 1999](#) taught children to read using the letter-sound rules for rime segments in words (i.e. the string of letters that follow an onset phoneme; e.g. w (onset) ine (rime)). [McArthur 2015a](#) and [McArthur 2015b](#) taught children how to read using computer programs that trained the pairings of graphemes (letter units) to phonemes (speech sounds) within the context of letter units, within parts of syllables, or within parts of regular words (note: children were not trained to read words per se).

Phonics and phoneme awareness

Seven of the 14 included studies trained poor readers with a programme that focused on training phoneme awareness as well as on training phonics-based reading skills ([Blythe 2006](#); [Ford 2009](#); [Hurford 1994](#); [Hurry 2007](#); [Lovett 2000](#); [Savage 2003](#); [Savage 2005](#)). [Blythe 2006](#) trained phoneme awareness, letter-sound rules, and blending. [Ford 2009](#) trained phonemic awareness and decoding multi-syllabic words using letter-sound rules. [Hurford 1994](#) trained various phoneme awareness skills (discrimination, segmentation, blending) with letters. [Hurry 2007](#) trained various phoneme awareness skills (alliteration, rhyme, boundary sounds, vowel sounds, digraph sounds (i.e. sounds associated with letter groups that make a single sound such as TH)), as well as using plastic letters to build words using letter-sound rules. [Lovett 2000](#) trained various phoneme awareness skills (segmentation, blending, rhyming) and used a special orthography (highlighting salient features of some letters) to teach letter-sound rules. [Savage 2003](#) and [Savage 2005](#) taught children to read using the letter-sound rules for phonemes (e.g. the letters C, S, and M) and rimes or rhymes (e.g. AT, as in CAT, SAT, MAT), and trained phoneme awareness for phonemes and rimes or rhymes.

Phonics and sight words

Two of the 14 included studies trained poor readers with a programme that focused on training children to read using letter-sound rules as they appear in whole words (e.g. the sound /sh/ was taught in the word /she/; [Chen 2014](#); [Lovett 1990](#)).

Training intensity

Less than two hours per week

Ten of the 14 included studies trained poor readers for less than two hours per week. Seven studies trained children between 60 and 90 minutes per week ([Barker 1995](#); [Blythe 2006](#); [Chen 2014](#); [Levy 1997](#); [Levy 1999](#); [Savage 2003](#); [Savage 2005](#)). Three studies trained children for, on average, 15 to 45 minutes per week ([Ford 2009](#); [Hurford 1994](#); [Hurry 2007](#)).

At least two hours per week

Four of the 14 included studies trained poor readers for four hours per week ([Lovett 1990](#); [Lovett 2000](#); [McArthur 2015a](#); [McArthur 2015b](#)).

Training duration

Less than three months

Twelve of the 14 included studies conducted their training for less than three months ([Barker 1995](#); [Blythe 2006](#); [Chen 2014](#); [Ford 2009](#); [Levy 1997](#); [Levy 1999](#); [Lovett 1990](#); [Lovett 2000](#); [McArthur 2015a](#); [McArthur 2015b](#); [Savage 2003](#); [Savage 2005](#)).

At least three months

Only two studies carried out training for over three months: [Hurford 1994](#) (five months) and [Hurry 2007](#) (seven months).

Training group size

One-to-one

Eight of the 14 included studies provided poor readers with one-to-one training by a reading professional (teacher, clinician, researcher) or computer ([Blythe 2006](#); [Ford 2009](#); [Hurford 1994](#); [Hurry 2007](#); [Levy 1997](#); [Levy 1999](#); [McArthur 2015a](#); [McArthur 2015b](#)).

Small group

Six of the 14 included studies trained poor readers in small groups comprising fewer than five trainees (Barker 1995; Chen 2014; Lovett 1990; Lovett 2000; Savage 2003; Savage 2005).

Training administrator

Human

Eight of the 14 included studies administered training primarily via a human, that is, researcher, teacher, or clinician (Chen 2014; Hurry 2007; Levy 1997; Levy 1999; Lovett 1990; Lovett 2000; Savage 2003; Savage 2005).

Computer

Six of the 14 included studies used computers as the primary training method (Barker 1995; Blythe 2006; Ford 2009; Hurford 1994; McArthur 2015a; McArthur 2015b).

Comparisons

All studies compared a phonics intervention to a control group that did either no training (i.e. treatment as usual; nine studies: Blythe 2006; Ford 2009; Hurry 2007; Hurford 1994; Levy 1997; McArthur 2015a; McArthur 2015b; Savage 2003; Savage 2005), or alternative training (five studies: Barker 1995; Chen 2014; Levy 1999; Lovett 1990; Lovett 2000).

Outcome measures

The tests used by each study to measure primary and secondary outcomes are outlined in the [Characteristics of included studies](#) table. They are summarised in [Table 1](#), and discussed below.

Primary outcomes

Mixed/regular word reading accuracy

Eleven of the 14 included studies measured mixed/regular word reading accuracy. In five studies, the tests were bespoke experimental tasks that presented readers with regular or irregular words (Levy 1997; Levy 1999; Lovett 1990; Lovett 2000; Savage 2003). Three studies used versions of the Word Identification subtest from the Woodcock-Johnson Reading Mastery Test (Barker 1995; Ford 2009; Hurford 1994). One study used the Wechsler Individual Achievement Test (Blythe 2006), one used the Word Reading Test from the British Ability Scales (Hurry 2007), and one used

the Group Reading Assessment and Diagnostic Evaluation Level 1 Version A: Word Recognition Assessment (Chen 2014).

Non-word reading accuracy

Ten of the 14 included studies tested non-word reading accuracy. Four studies used a non-word reading test from a version of the Woodcock-Johnson Reading Mastery Test (Barker 1995; Ford 2009; Hurford 1994; Lovett 2000), five studies used experimental non-word reading tests that were developed specifically for the study (Levy 1997; Levy 1999; McArthur 2015a; McArthur 2015b; Savage 2003), and one study used a non-word reading test from the Wechsler Individual Achievement Test (Blythe 2006).

Irregular word reading accuracy

Four of the 14 included studies tested irregular word reading accuracy through experimental irregular word reading tests that were developed for the study (Lovett 1990; Lovett 2000; McArthur 2015a; McArthur 2015b).

Mixed/regular word reading fluency

Four of the 14 included studies measured mixed/regular word reading fluency. Three studies used the Sight Word test from the Test of Word Reading Efficiency (Ford 2009; McArthur 2015a; McArthur 2015b). A fourth study used two experimental tests of regular and irregular words that were designed specifically for the study (Lovett 1990). For the meta-analysis in this review, we calculated the mean effect sizes of the two outcomes used in Lovett 1990 using the procedures outlined in [Unit of analysis issues](#).

Non-word reading fluency

Four of the 14 included studies tested non-word reading fluency using the Phonemic Decoding test from the Test of Word Reading Efficiency (Ford 2009; Lovett 1990; McArthur 2015a; McArthur 2015b).

Reading comprehension

Five of the 14 included studies tested reading comprehension. One study used the Neale Analysis of Reading Ability (Hurry 2007), one used the Wechsler Individual Achievement Test (Blythe 2006), and one used the Gates-MacGinitie Reading Test (Ford 2009). Two studies used the Test of Everyday Reading Comprehension (McArthur 2015a; McArthur 2015b).

Spelling

Three of the 14 included studies tested mixed/regular or irregular word spelling (Chen 2014; Lovett 1990; Savage 2003). Lovett 1990 tested spelling with separate mixed/regular and irregular spelling tests that were designed specifically for the study. For the meta-analysis in this review, we calculated the mean effect sizes of the two outcomes used in Lovett 1990 using the procedures outlined in [Unit of analysis issues](#). Chen 2014 and Savage 2003 used experimental spelling tests that were developed specifically for each study.

Secondary outcomes

Letter-sound knowledge

Three of the 14 included studies tested letter-sound knowledge. This was unexpected since letter-sound knowledge is the focus of phonics training. The three studies tested letter-sound knowledge using experimental tasks designed specifically for the study (Lovett 1990; Savage 2003; Savage 2005).

Phonological output

Four of the 14 included studies tested phonological output. Three studies used experimental tasks designed specifically for the study (Barker 1995; Savage 2003; Savage 2005), and one study used the Goldman Fristoe Woodcock Sound Analysis test (Lovett 2000).

Funding

Of the 14 studies included in this review, eight stated that they were supported by funding organisations: Qualifications and Curriculum Authority (Hurry 2007), Ontario Mental Health Foundation (Levy 1997; Lovett 2000), Social Sciences and Humanities Research Council of Canada (Levy 1999; Lovett 1990), Velleman Foundation (Lovett 2000), National Institute of Health and Child Development (Lovett 2000), National Health and Medical Research Council (McArthur 2015a; McArthur 2015b), Australian Research Council (McArthur 2015a; McArthur 2015b), JJ Trust (Savage 2003), Helen Arkell Dyslexia Association (Savage 2003), and McGill University (Savage 2003).

Excluded studies

In the [Characteristics of excluded studies](#) table, we listed studies that reading researchers might expect to be included in this review but were excluded because they failed to meet our review criteria ([Criteria for considering studies for this review](#)). We excluded 29 studies because interventions included phonics training plus two or more other skills such as text reading, phoneme

awareness, and reading comprehension (Foorman 1997; Foorman 1998; Gillon 1997; Goldstein 2017; Gorard 2015; Hatcher 1994; Hatcher 2006; Jeffes 2016; King 2015; Lovett 1988; Lovett 1989; Lovett 2012; Merrell 2015; Metsala 2017; Munro 2017; Olson 1997; Rashotte 2001; Schlesinger 2017; Seiler 2018; Steacy 2016; Storey 2017; Torgesen 1999a; Torgesen 2006; Vellutino 1986; Vellutino 1987; Vellutino 1996; Wheldall 2017; Wise 1997; Wise 1999). Thirteen other studies did not use randomisation, quasi-randomisation, or minimisation (Gillon 2000; Gillon 2002); did not assess reading at pre- and post-training (Torgesen 1997); used participants who did not meet our inclusion criteria (Arnold 2016; Bhide 2013; Christodoulou 2017; Dubois 2014; Lovett 1994; Savage 2018; Schaars 2017; Van Gorp 2017); did not include a control group that was untrained or did non-phonics alternative training (Alexander 1991); or was a review paper (Olson 1992). We excluded six studies for multiple reasons: four studies did not meet the criteria for phonics training or did not include a control group that was untrained or did non-phonics alternative training (Berninger 2013; Torgesen 2001; Wise 1995; Wise 2000); and in two studies neither the intervention nor the participants met our inclusion criteria (Aboud 2018; Messer 2018).

Risk of bias in included studies

Below, and in [Figure 1](#), we provided a summary of the results of our 'Risk of bias' assessment for each included study. Further details can be found in the 'Risk of bias' tables (see [Characteristics of included studies](#) tables).

Ten studies did not describe their random sequence generation (Barker 1995; Blythe 2006; Hurford 1994; Hurry 2007; Levy 1997; Levy 1999; Lovett 1990; Lovett 2000; Savage 2003; Savage 2005), and 11 studies did not provide information about allocation concealment and blinding (Barker 1995; Blythe 2006; Ford 2009; Hurford 1994; Hurry 2007; Levy 1997; Levy 1999; Lovett 1990; Lovett 2000; Savage 2003; Savage 2005). We contacted the authors of these studies and all supplied further information regarding these 'Risk of bias' factors (see directly below and the 'Risk of bias' tables for each study for more information).

Allocation

Random sequence generation

Information provided in publications and from personal communications with study authors indicated that all studies allocated participants to groups using randomisation (Barker 1995; Blythe 2006; Chen 2014; Ford 2009; Hurford 1994; Hurry 2007; Levy 1997; Levy 1999; Lovett 1990; Lovett 2000; Savage 2003; Savage 2005), quasi-randomisation (McArthur 2015a), or minimisation (McArthur 2015b). See [Characteristics of included studies](#) table for details. The study that used quasi-randomisation provided evidence for why risk of bias was low (see [Characteristics](#)

of included studies table). Thus, we rated all studies at low risk of bias on this domain.

Allocation concealment

All studies used central allocation of participants to groups so personnel could not have foreseen assignment due to groups. Therefore, we rated all studies at low risk of bias on this domain.

Blinding

Participants and personnel

It is difficult to absolutely ensure blinding of personnel that deliver any cognitive treatment, which are often delivered by humans who must be aware of what they are doing. Blinding of participants in cognitive reading treatment trials is of less concern since participants (mostly children) do not have the expertise to discern the nature of the experimental or control intervention - if, indeed, they are aware a control intervention exists. Thus, degree of performance bias in the current review was primarily driven by how a study tackled the blinding of personnel. Seven studies employed methods or procedures that explicitly addressed personnel blinding (Blythe 2006; Levy 1997; Levy 1999; McArthur 2015a; McArthur 2015b; Savage 2003; Savage 2005). We judged these studies at low risk of performance bias. The seven remaining studies did not report explicit attempts to minimise performance bias and hence we judged them at unclear risk of performance bias (Barker 1995; Chen 2014; Ford 2009; Hurford 1994; Hurry 2007; Lovett 1990; Lovett 2000).

Outcome assessment

Concerns about blinding of outcome assessment in reading trials are mitigated by the fact that such trials use objective tests of literacy-related skills that are explicitly designed to avoid assessor bias via standardised administration and scoring procedures. In this review, nine studies employed such tests and made explicit attempts to address blinding of outcome assessment (Ford 2009; Hurford 1994; Hurry 2007; Lovett 1990; Lovett 2000; McArthur 2015a; McArthur 2015b; Savage 2003; Savage 2005). We judged these studies at low risk of detection bias. Two studies used objective literacy-related tests but did not report explicit attempts to minimise blinding of outcome assessment (Barker 1995; Chen 2014), while the three remaining studies used objective literacy-related tests and reported that it did not make an explicit attempt to minimise blinding of outcome assessment (Blythe 2006; Levy 1997; Levy 1999). We rated these five studies at unclear risk of bias.

Incomplete outcome data

Five of the 14 included studies indicated that there was no attrition across the study, so we judged these at low risk of attrition bias (Blythe 2006; Chen 2014; Levy 1997; Levy 1999; Lovett 2000). Eight studies reported minor attrition across the study, with groups similarly affected, so we judged these at low risk of attrition bias also (Ford 2009; Hurford 1994; Hurry 2007; Lovett 1990; McArthur 2015a; McArthur 2015b; Savage 2003; Savage 2005). One study did not provide any information about incomplete outcome data, so we judged it at unclear risk of attrition bias (Barker 1995).

Selective reporting

For all but one study, there were no apparent missing literacy tests (Chen 2014); Chen 2014 did not provide post-test data for a single secondary outcome - blinding. Nevertheless, the absence of review protocols, or explicit statements by studies that no tests had been excluded from analysis or publication, meant that we had to rate all studies as unclear for selective reporting.

Other potential sources of bias

No study reported any other potential sources of bias, and hence we rated all studies at low risk of other potential sources of bias.

Effects of interventions

See: [Summary of findings for the main comparison Phonics training versus control \(no training or alternative training\) for English-speaking poor readers](#)

The primary aim of this review was to measure the effect of phonics training on literacy-related skills in English-speaking poor readers. To this end, we calculated the effects of phonics training on seven primary and two secondary outcomes (see below). A summary of the statistics, including GRADE quality ratings, can be found in [Summary of findings for the main comparison](#) and [Table 1](#). A summary of the tests used to measure the outcomes can be found in [Table 2](#).

Primary outcomes

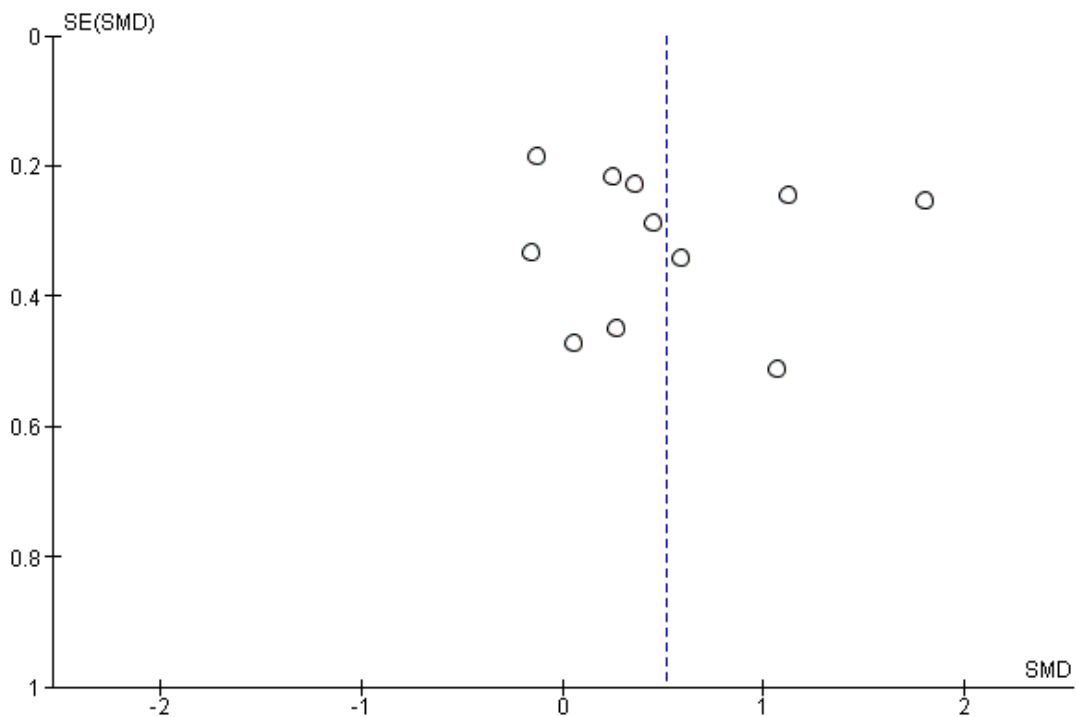
Mixed/regular word reading accuracy

Eleven of the 14 studies (701 participants) tested the effect of phonics on mixed/regular word reading accuracy (Barker 1995; Blythe 2006; Chen 2014; Ford 2009; Hurford 1994; Hurry 2007; Levy 1997; Levy 1999; Lovett 1990; Lovett 2000; Savage 2003). Three studies used multiple mixed/regular reading tests (Barker 1995; Lovett 1990; Lovett 2000). We dealt with repeated measures of the same outcome as outlined in the [Unit of analysis issues](#)

section, and further explained in the respective [Characteristics of included studies](#) table. Heterogeneity for this outcome was considerable, exceeding 70% ($\text{Chi}^2 = 52.11$; $P < 0.001$; $I^2 = 81\%$). We wondered if the large I^2 value was due to the atypical negative effect found by [Barker 1995](#) (SMD -0.16) and the unusually large effect found by [Levy 1999](#) (SMD 1.80). However, following the three steps outlined in [Assessment of heterogeneity](#), we concluded that there was no reason to adjust or exclude the data from any studies, and hence we compared a fixed-effect and random-effects meta-analysis for this outcome. The results were similar (see [Table 2](#)), so we focused on those from the random-effects model, which adjusted estimates to incorporate heterogeneity ([Deeks 2011](#)). The SMD was 0.51 (95% CI 0.13 to 0.90; $Z = 2.59$; $P = 0.01$;

[Analysis 1.1](#)). The GRADE rating for this moderate effect was low. According to criteria outlined by [Ryan 2016](#), this means that phonics training in English-speaking poor readers may improve mixed/regular word reading accuracy. More data are required to increase the precision of the data and certainty of this effect. We drew a funnel plot to explore reporting bias for the one outcome that had data from more than 10 studies and did not have similar standard errors for their effect sizes (mixed/regular word reading accuracy). The plot showed that studies with the least power and imprecision (at the bottom of the graph) did not scatter more widely than those at the top. This suggested an absence of reporting bias. See [Figure 3](#).

Figure 3. Funnel plot of comparison: I Treatment versus control random-effects model, outcome: I.1 Mixed/regular word reading accuracy.



Non-word reading accuracy

Ten of the 14 studies (682 participants) used eight different measures to test the effect of phonics on non-word reading accuracy ([Barker 1995](#); [Blythe 2006](#); [Ford 2009](#); [Hurford 1994](#); [Levy 1997](#);

[Levy 1999](#); [Lovett 2000](#); [McArthur 2015a](#); [McArthur 2015b](#); [Savage 2003](#)). Heterogeneity for this outcome exceeded 70% ($\text{Chi}^2 = 50.72$; $P < 0.001$; $I^2 = 82\%$). We wondered if it was due to an atypical negative effect found by [Barker 1995](#) (SMD -0.50). Following the steps outlined in [Assessment of heterogeneity](#), we

determined not to adjust or exclude the data from any studies, and hence we compared fixed-effect and random-effects meta-analyses. The results were similar (see [Table 2](#)), so we focused on those from the random-effects model.

The SMD was 0.67 (95% CI 0.26 to 1.07; $Z = 3.24$; $P = 0.001$; [Analysis 1.2](#)). The GRADE rating for this moderate effect was low. According to [Ryan 2016](#)'s criteria, phonics training in English-speaking poor readers may improve non-word reading accuracy. More data are required to increase the precision of the data and certainty of this effect.

Irregular word reading accuracy

Four of the 14 studies (294 participants) tested the effect of phonics training on irregular word reading accuracy in poor readers ([Lovett 1990](#); [Lovett 2000](#); [McArthur 2015a](#); [McArthur 2015b](#)). We dealt with repeated measures of the same outcome in [Lovett 1990](#) as outlined in the [Unit of analysis issues](#) section, and further explained in the [Characteristics of included studies](#) table. Heterogeneity for this outcome exceeded 70% ($\text{Chi}^2 = 14.41$; $P = 0.002$; $I^2 = 79\%$). As per the steps outlined in [Assessment of heterogeneity](#), we identified no reason to adjust or exclude the data from any studies, and hence we compared fixed-effect and random-effects meta-analyses. The results were similar for the two analyses (see [Table 2](#)), so we focus on those from the random-effects model.

The SMD was 0.84 (95% CI 0.30 to 1.39; $Z = 3.04$; $P = 0.002$; [Analysis 1.3](#)). The GRADE rating for this large effect was moderate. According to [Ryan 2016](#), phonics training in English-speaking poor readers probably improves irregular word reading accuracy. More data are required to increase the precision of the data and certainty of this effect.

Mixed/regular word reading fluency

Four of the 14 studies (224 participants) tested the effect of phonics on mixed/regular word reading fluency ([Ford 2009](#); [Lovett 1990](#); [McArthur 2015a](#); [McArthur 2015b](#)). We dealt with repeated measures of the same outcome in [Lovett 1990](#) as outlined in the [Unit of analysis issues](#) section, and further explained in the respective [Characteristics of included studies](#) table. We dealt with inverted scale issues in [Lovett 1990](#) (i.e. high scores represented poorer performance - in contrast to the three other studies) as outlined in the [Characteristics of included studies](#) table. Heterogeneity for this outcome was low ($\text{Chi}^2 = 2.20$; $P = 0.53$; $I^2 = 0\%$).

The SMD was 0.45 (95% CI 0.19 to 0.72; $Z = 3.33$; $P < 0.001$; [Analysis 1.4](#)). The GRADE rating for this moderate effect was moderate. Thus, phonics training in English-speaking poor readers probably improves mixed/regular word reading fluency ([Ryan 2016](#)); however, more data are required to increase the precision of the data and the certainty of this effect.

Non-word reading fluency

Three of the 14 studies (188 participants) tested the effect of phonics on non-word reading fluency ([Ford 2009](#); [McArthur 2015a](#); [McArthur 2015b](#)). Heterogeneity for this outcome was low ($\text{Chi}^2 = 0.02$; $P = 0.99$; $I^2 = 0\%$).

The SMD was 0.39 (95% CI 0.10 to 0.68; $Z = 2.63$; $P = 0.009$; [Analysis 1.5](#)). The GRADE rating for this moderate effect was moderate, suggesting that phonics training in English-speaking poor readers probably improves non-word reading fluency ([Ryan 2016](#)). However, more data are required to increase the precision of the data and the certainty of this effect.

Reading comprehension

Five of the 14 studies (343 participants) tested the effect of phonics on reading comprehension ([Blythe 2006](#); [Ford 2009](#); [Hurry 2007](#); [McArthur 2015a](#); [McArthur 2015b](#)). Heterogeneity for this outcome was moderate ($\text{Chi}^2 = 8.45$; $P = 0.08$; $I^2 = 53\%$).

The SMD was 0.28 (95% CI -0.07 to 0.62; $Z = 1.54$; $P = 0.12$; [Analysis 1.6](#)). The GRADE rating for this small effect was low, which means that phonics training in English-speaking poor readers may slightly improve poor reading comprehension ([Ryan 2016](#)). More data are required to increase the precision of the data and the certainty of this effect.

Spelling

Three of the 14 studies (158 participants) tested the effect of phonics on spelling words ([Chen 2014](#); [Lovett 1990](#); [Savage 2005](#)). We dealt with repeated measures of the same outcome in [Lovett 1990](#) as outlined in the [Unit of analysis issues](#) section, and further explained in the respective [Characteristics of included studies](#) table. Heterogeneity for this outcome was moderate ($\text{Chi}^2 = 3.89$; $P = 0.14$; $I^2 = 49\%$).

The SMD was 0.47 (95% CI -0.07 to 1.01; $Z = 1.72$; $P = 0.09$; [Analysis 1.7](#)). The GRADE rating for this moderate effect was low, meaning that phonics training in English-speaking poor readers may improve poor spelling ([Ryan 2016](#)). More data is required to increase the precision of the data and the certainty of this effect.

Secondary outcomes

Letter-sound knowledge

Three of the 14 studies (192 participants) tested the effect of phonics on letter-sound knowledge ([Lovett 1990](#); [Savage 2003](#); [Savage 2005](#)). We dealt with repeated measures of the same outcome in [Lovett 1990](#) as outlined in the [Unit of analysis issues](#) section, and further explained in the respective [Characteristics of included studies](#) table. The heterogeneity for this outcome was low ($\text{Chi}^2 = 0.11$; $P = 0.95$; $I^2 = 0\%$).

The SMD was 0.35 (95% CI 0.04 to 0.65; $Z = 2.22$; $P = 0.03$; [Analysis 1.8](#)). The GRADE rating for this moderate effect was low. According to [Ryan 2016](#), this means that phonics training in English-speaking poor readers may improve letter-sound knowledge. More data are required to increase the precision of the data and the certainty of this effect.

Phonological output

Four of the 14 studies (280 participants) tested the effect of phonics on phonological output ([Barker 1995](#); [Lovett 2000](#); [Savage 2003](#); [Savage 2005](#)). Following the steps in [Assessment of heterogeneity](#), we identified no reason to adjust or exclude the data from any studies, and hence we compared fixed-effect and random-effects meta-analyses. The results were similar for the two analyses (see [Table 2](#)), so we focused on those from the random-effects model. The SMD was 0.38 (95% CI -0.04 to 0.80; $Z = 1.77$; $P = 0.08$; [Analysis 1.9](#)). The GRADE rating for this moderate effect was low. According to [Ryan 2016](#), this means that phonics training in English-speaking poor readers may improve phonological output. More data are required to increase the precision of the data and the certainty of this effect.

Subgroup analyses

The secondary aim of this review was to explore the impact of moderating factors on the efficacy of phonics training in poor readers (see [Analysis 2.1](#); [Analysis 2.2](#)). A summary of the statistics can be found in [Table 5](#), which shows that: no subgroup analysis included more than nine studies (most comprised only two to seven studies); and the heterogeneity of data with most subgroups was high (i.e. I^2 greater than 70%). Therefore, we concluded that there were not enough reliable data to make confident conclusions from these subgroup analysis at this time.

Sensitivity analyses

In addition to the sensitivity analyses that compared the results from a fixed-effect meta-analysis with those from a random-effects meta-analyses for outcomes with high heterogeneity (reported above), we conducted the two following sensitivity analyses.

Random sequence generation

The combined information from publications and personal communications (see [Included studies](#) and [Table 6](#)) indicated that almost all studies were controlled trials that used randomisation or minimisation. We were unsure about [Hurford 1994](#) due to the discrepancy between the published report and personal contact with the author, so we included the study but undertook sensitivity analyses to determine the impact of removing it. [Hurford 1994](#) contributed data to just two outcomes: mixed/regular word reading accuracy and non-word reading accuracy. The SMD for

mixed/regular word reading accuracy with and without [Hurford 1994](#) were 0.51 (95% CI 0.13 to 0.90; $Z = 2.59$; $P = 0.01$; 11 studies, 701 participants; [Analysis 1.1](#)) and 0.52 (95% CI 0.09 to 0.95; $Z = 2.38$; $P = 0.02$; 10 studies, 651 participants; [Analysis 4.1](#)), respectively. The SMD for non-word reading accuracy with and without [Hurford 1994](#) were 0.67 (95% CI 0.26 to 1.07; $Z = 3.24$; $P = 0.001$; 10 studies, 682 participants; [Analysis 1.2](#)) and 0.69 (95% CI 0.24 to 1.14; $Z = 3.03$; $P = 0.002$; 9 studies, 632 participants; [Analysis 4.2](#)), respectively. These very similar outcomes suggest that the unclear random allocation for [Hurford 1994](#) did not have undue influence on the overall outcomes.

Group size

We undertook a second sensitivity analysis to determine the influence of three relatively small studies on the outcomes (i.e. 10 or fewer participants in the experimental and control groups): [Blythe 2006](#) ($n = 10$), [Chen 2014](#) ($n = 9$), and [Ford 2009](#) ($n = 9$). The SMD for the outcomes tested by these studies were very similar to those for the full study set: mixed/regular word reading accuracy: 0.53 (95% CI 0.07 to 1.00; $Z = 2.26$; $P = 0.02$; 8 studies, 654 participants; [Analysis 5.1](#)); non-word reading accuracy: 0.66 (95% CI 0.20 to 1.11; $Z = 2.82$; $P = 0.005$; 8 studies, 644 participants; [Analysis 5.2](#)); irregular word reading accuracy: 0.84 (95% CI 0.30 to 1.39; $Z = 3.03$; $P = 0.002$; 4 studies, 294 participants; [Analysis 5.3](#)), mixed/regular word reading fluency: 0.50 (95% CI 0.22 to 0.78; $Z = 3.53$; $P < 0.001$; 3 studies, 206 participants; [Analysis 5.4](#)), non-word reading fluency: 0.39 (95% CI 0.08 to 0.69; $Z = 2.50$; $P = 0.01$; 2 studies, 170 participants; [Analysis 5.5](#)), reading comprehension: 0.25 (95% CI -0.15 to 0.64; $Z = 1.22$; $P = 0.22$; 3 studies; 305 participants; [Analysis 5.6](#)), and spelling: 0.36 (95% CI -0.27 to 0.99; $Z = 1.12$; $P = 0.26$; 2 studies, 140 participants; [Analysis 5.7](#)). Thus, small studies did not appear to have undue influence on the overall outcomes.

DISCUSSION

Summary of main results

We included 14 studies with 923 participants in this review; three studies were new to this update ([Chen 2014](#); [McArthur 2015a](#); [McArthur 2015b](#)). A meta-analysis of the data revealed that phonics training in English-speaking poor readers probably improved irregular word reading accuracy, mixed/regular word reading fluency, and non-word reading fluency. It may improve mixed/regular word reading accuracy, non-word reading accuracy, spelling, letter-sound knowledge, and phonological output. And it may slightly improve reading comprehension. The positive effects of phonics training on all outcomes indicated that phonics did not harm literacy-related skills in English-speaking poor readers. The quality of evidence provided by the studies that generated these

effect sizes was moderate or low. The wide CIs around the SMD for each outcome raised concern about the precision of the data, and highlighted the need for studies that assess the effect of phonics training on more homogeneous groups of poor readers who are known to have problems in their phonics skills that match the types of phonics training included in a programme.

This review conducted a series of subgroup analyses to determine if phonics training in English-speaking poor readers is modulated by training type, training intensity, training duration, or training group size. A lack of studies, and high heterogeneity in results, meant that we could not draw conclusions with any confidence. Many more studies are needed to determine if any of these factors modulate the effect of phonics training in English-speaking poor readers.

Overall completeness and applicability of evidence

The outcomes of the 14 studies in this review appear applicable to English-speaking poor readers in the general population for at least five reasons. First, the fact that many of the studies were published since 2003 indicates that the findings are applicable to poor word readers in current times (Blythe 2006; Chen 2014; Ford 2009; Hurry 2007; McArthur 2015a; McArthur 2015b; Savage 2003; Savage 2005). Second, a similar number of studies were done in each of the major English-speaking countries in the world; specifically, studies were done in Canada (five studies), the USA (three studies), the UK (three studies), and Australia (three studies). Third, research has established that poor reading is not restricted to a particular culture or SES. The studies in this review recruited samples with a variety of ethnic backgrounds and SES, which is representative of English-speaking poor readers in the general population. Fourth, the studies included similar numbers of males and females. There is a popular perception that more males than females are poor readers. This view has arisen from recruitment bias: people are more likely to notice poor reading in boys than girls, possibly because boys are more likely to misbehave when they are frustrated or bored. Studies minimising recruitment bias have found about equal proportions of male and female poor readers (Shaywitz 2001). Thus, by recruiting similar numbers of males and females, the studies included in this review represented the proportion of males and females with poor reading in the general population. The fifth reason related to IQ. Most poor readers from the studies included in this review had IQ scores within or above the mean range. This reflects the type of poor reader who gains the most attention in society (i.e. those with poor reading despite average intelligence). As mentioned in the [Background](#) section, there is growing evidence that IQ is not predictive of poor reading or response to intervention. Thus, the outcomes of this review are applicable to poor readers with various levels of IQ. It is noteworthy that all but one study tested children, and so the results of this review are more directly applicable to children than

adults. There is currently no evidence to suggest that adults with poor reading respond differently to phonics training than children with poor reading, but the number of studies addressing this issue is limited. It is also noteworthy that only three studies included letter-sound knowledge as an outcome measure. This is surprising given that phonics training focuses on this skill. Future studies should include letter-sound knowledge measures to ensure a more complete understanding of the effects of phonics on poor readers.

Quality of the evidence

There are at least five factors that have the potential to affect the quality of evidence in this review.

First, there is risk of bias. As illustrated by the 'Risk of bias' table (Figure 1), all studies had a low-risk judgement for the majority of the seven biases assessed in this review (see Figure 1).

Second factor was quality of evidence. According to GRADE (Schünemann 2011a), randomised and pseudo randomised trials - the only type of trial included in this review - are initially rated as high quality, with quality subsequently downgraded for a number of factors (see Table 6). The quality of evidence was moderate for the three primary outcomes of irregular word reading accuracy, mixed regular/regular word reading fluency, and non-word reading fluency, but low for the four primary outcomes (mixed/regular word reading accuracy, non-word reading accuracy, reading comprehension, and spelling). For all primary outcomes, the quality of evidence was limited by imprecision, with wide CIs around the SMDs for each. This is unsurprising given that poor readers are a heterogeneous population. There is a great need for future studies to closer match the nature of the phonics problems in a sample and the type of phonics training delivered. This will reduce the heterogeneity of response to phonics training and hence improve the precision of results.

Third, there is variability in the amount of data used to calculate effects for each outcome. While the effects for mixed/regular word reading accuracy and non-word reading accuracy were calculated from 10 to 11 studies, the effects for the remaining nine outcome measures were calculated from three to five studies.

Fourth, there was the chance that some training studies exposed participants in a treatment group - but not a control group - to content that is included in the outcomes. While it is possible that some phonics training programmes may expose children to words, or parts of words, that may be included in the post-tests, phonics training programmes typically use a wide range of constantly changing stimuli to teach children the letter-sound rules, rather than repeatedly using the same content (i.e. specific words or non-words). Since phonics training typically focuses on repeatedly training rules, rather than specific content, the effect of content exposure during training should be minimal in typical phonics training studies.

Fifth, the quality of evidence may be affected by publication bias. As outlined in the methods, we planned to use funnel plots to

explore reporting bias for any outcome that had data from more than 10 studies which did not have similar standard errors for their effect sizes (Sterne 2011). In this review, only one outcome had data from more than 10 studies (mixed/regular word reading accuracy). The funnel plot for this outcome, which is shown in Figure 3, indicated that studies with the least power and imprecision (at the bottom of the graph) did not scatter more widely than those at the top. This suggested an absence of bias against publishing small studies with non-significant effects (in which case there would be a clear gap in the bottom left of the graph), or towards publishing studies based on P values alone (in which case, the plot would have more studies at the left and right sides of the graph than in the middle (Sterne 2011). Thus, publication bias did not appear to account for the heterogeneity for the mixed/regular word reading accuracy outcome at least.

Potential biases in the review process

The various analyses conducted in this review suggested that potential biases were minimal for seven reasons. First, almost all studies had low risk of bias for random sequence generation, incomplete outcome data, and selective reporting. The majority also had low or unclear risk of bias for allocation concealment, blinding of outcome assessment, and blinding of personnel and participants. Second, excessive heterogeneity applied to just three outcomes, with an analysis revealing no systematic explanation for this variance. Third, a funnel plot of the mixed/regular word reading accuracy outcome suggested no evidence of publication bias, bias introduced by using P values, or bias owing to outliers. Fourth, a comparison of effects using fixed- and random-effects analyses revealed very similar outcomes, suggesting a degree of statistical reliability. Fifth, two sensitivity analyses produced very similar results to the primary analysis. Sixth, the quality of evidence for most primary outcomes was moderate. Seventh, we ensured that review authors who were also authors of the included studies did not assess the eligibility of these studies for inclusion and did not extract the data.

Agreements and disagreements with other studies or reviews

There are two previous meta-analyses that are highly relevant to this review. The National Reading Panel in the USA found small-to-moderate effects of phonics on the reading skills of poor readers (Ehri 2001). In line with this, the current review found moderate effects of phonics training on mixed/regular word reading accuracy, non-word reading accuracy, mixed/regular word reading fluency, non-word reading accuracy, spelling, letter-sound knowledge and phonological output; and a small effect on reading comprehension. Interestingly, this review found large effects of phonics

training on irregular word reading accuracy in English-speaking poor readers.

A likely explanation for any slight discord in the results between the two reviews was the different criteria used for study inclusion. In the current study, we were interested in the specific effect of phonics training. Ideally, we would have only included studies that used 'pure' phonics training programmes (i.e. programmes that only taught reading via phonics-based reading skills). However, prior to doing this review, we suspected pure phonics training studies might be rare. Thus, our criteria for phonics training included programmes that trained phonics alone, or trained phonics plus one other reading-related skill (sight word reading, phoneme awareness). The National Reading Panel in the USA did not use such strict criteria, and so included many more studies that used programmes that trained at least two other reading skills in addition to phonics (Ehri 2001). As discussed above, the outcomes of such complex phonics programmes are difficult to interpret because reading gains could stem from phonics training, non-phonics training, or an interaction between the two. The fact that the current review found moderate and large effects for some outcomes suggests that the inclusion of non-phonics training in complex phonics programmes may weaken training effects for some reading-related outcomes - perhaps because less time is dedicated to phonics training per se.

A second previous meta-analysis, conducted by Suggate 2010, found a moderate effect size of phonics training on reading skills, prereading skills, and comprehension skills in children who were struggling readers. Suggate's criteria for phonics training were quite similar to the current review, and Suggate's criteria for struggling readers were similar to our criteria for poor readers (see [Criteria for considering studies for this review](#)). This may explain why our moderate effects reflect those of Suggate, and why Suggate identified a similar number of relevant phonics training studies (13) in struggling readers in the first seven years of school. However, unlike the current review, Suggate 2010 focused on children and did not include unpublished studies. Thus, the slightly different outcomes of the two studies could be explained by different study sets.

A third previous meta-analysis - carried out by Galushka 2014 - found a small but significant effect of phonics training on reading skills of children and adolescents with poor reading. As mentioned above, this review included a variety of phonics training, many of which trained phonics in conjunction with other skills. Thus, as per Ehri 2001, the inclusion of non-phonics training in such multi-faceted phonics programmes may have weakened the effect of phonics per se on reading. In addition, Galushka 2014 calculated the mean effect of phonics across different reading outcome measures. In the current review, with the exception of the irregular word reading result (based on only four studies), there is evidence that phonics had larger effects on reading outcomes that depended directly upon phonics-related reading skills (e.g. non-word reading accuracy, mixed/regular word reading accuracy) than

those that also depended upon other cognitive skills (e.g. reading comprehension, phonological output).

AUTHORS' CONCLUSIONS

Implications for practice

The outcomes of this review suggest that phonics training is probably effective for treating poor reading fluency for regular words and non-words. It may be effective for treating poor reading accuracy for regular words, non-words, and irregular words, as well as poor spelling, letter-sound knowledge, and phonological output. It may be slightly effective for treating poor reading comprehension. The positive effects of phonics training on all outcomes suggests that phonics training is not harmful for English-speaking poor readers. These findings suggest that phonics training is an appropriate treatment of choice to improve certain literacy-related skills in poor readers.

Implications for research

There is a widely held belief that phonics training is the best way to treat poor reading, yet we found only 14 studies that examined the effect of phonics training specifically in English-speaking poor readers. That is, only 14 studies have tested the effect of a training programme that used either phonics alone, or phonics plus one other literacy-related skill. More studies are needed to further improve our confidence about the strength and extent of the specific effect of phonics training in English-speaking poor readers.

More studies are also needed to assess the effect of phonics training on skills beyond word reading accuracy. For example, only three studies tested non-word reading fluency, and only three studies measured letter-sound knowledge - a surprising finding given this is the primary focus of phonics training. Future studies should include a more comprehensive range of reading outcomes to further understand the underlying cognitive processes that are influenced by phonics training in poor readers.

More research is needed to understand the effect that moderator variables - such as training type, training intensity, training dura-

tion, training group size, and training administrator - have on the effectiveness of phonics training for poor readers. In this review, we attempted to address these issues via the subgroup analyses for each outcome. However, the number of studies contributing to each subgroup was too small, and the heterogeneity of data for the majority of outcomes too high, to draw any conclusions with confidence.

A surprising finding of this review was that phonics training had its largest effect on irregular word reading accuracy. According to most models of word reading, phonics training should have its largest effect on the ability to read regular words or non-words (Coltheart 2001; Harm 1999).

Finally, future studies of phonics training for poor readers need to report more explicitly their methods for generating allocation sequences and concealment. While double blinding is difficult to guarantee in cognitive treatment trials, few studies explained how they at least attempted to instigate double blinding. Thus, future RCTs of phonics programmes need to explain the methods of their randomised controlled trials in more detail. The CONSORT 2010 guidelines may prove useful in this respect (Shultz 2010).

ACKNOWLEDGEMENTS

We are extremely and eternally grateful for the help provided by the Cochrane Developmental, Psychosocial and Learning Problems editors on this updated review, as well as on the original review. In particular, we would like to thank Geraldine Macdonald (updated and previous review), Joanne Duffield (updated review), Margaret Anderson (updated and previous review), Laura MacDonald (previous review), and Nuala Livingstone (previous review). We would also like to thank the statistician and external reviewers of the protocol and review, including this updated version; and we would like to acknowledge the efforts of three contributors to the previous version of the review (Pip Eve, Kristy Jones, and Linda Larsen). Finally, we would like to thank the authors of the studies included in this review for responding so willingly to repeated requests for information.

REFERENCES

References to studies included in this review

- Barker 1995** {*published data only (unpublished sought but not used)*}
 Barker TA, Torgese JK. An evaluation of computer-assisted instruction in phonological awareness with below average readers. *Journal of Educational Computing Research* 1995;**13**(1):89–103. DOI: 10.2190/TH3M-BTP7-JEJ5-JFNJ
- Blythe 2006** {*published data only (unpublished sought but not used)*}
 Blythe J. Cochrane review data request [pers comm]. Email to: P Eve 4 October 2011.
 * Blythe JM. Computer-based phonological skills training for primary students with mild to moderate dyslexia - a pilot study. *Australian Journal of Educational and Developmental Psychology* 2006;**6**:39–49. EJ815611
 Eve P. Cochrane review data request [pers comm]. Email to: J Blythe 14 September 2011.
- Chen 2014** {*published data only*}
 Chen V. Cochrane review - additional information [pers comm]. Email to: D Francis 11 July 2017.
 * Chen V, Savage RS. Corrigendum to the 'Evidence for a simplicity principle: teaching common complex grapheme-to-phonemes improves reading and motivation in at-risk readers' *Journal of Research in Reading*, 2014;**37**(2): 196–214. *Journal of Research in Reading* 2016;**39**(1):126–31. DOI: 10.1111/1467-9817.12068
 Chen V, Savage RS. Evidence for a simplicity principle: teaching common complex grapheme-to-phonemes improves reading and motivation in at-risk readers. *Journal of Research in Reading* 2014;**37**(2):196–214. DOI: 10.1111/1467-9817.12022
- Ford 2009** {*unpublished data only*}
 Ford C. *The effect of the backward-chaining method of decoding with computer-assisted instruction on the reading skills of struggling adolescent readers [Doctoral thesis]*. DeKalb (IL): Northern Illinois University, 2009.
- Hurford 1994** {*published data only (unpublished sought but not used)*}
 Hurford D. Cochrane review data request [pers comm]. Email to: P Eve 4 February 2012.
 * Hurford DP, Johnston M, Nepote P, Hampton S, Moore S, Neal J, et al. Early identification and remediation of phonological-processing deficits in first-grade children at risk for reading disabilities. *Journal of Learning Disabilities* 1994;**27**(10):647–59. DOI: 10.1177/002221949402701005; PUBMED: 7844481
- Hurry 2007** {*published data only (unpublished sought but not used)*}
 Eve P. Cochrane review data request [pers comm]. Email to: D Hurry 14 September 2011.
 Hurry J. Cochrane review data request [pers comm]. Email to: P Eve 16 February 2012.
 Hurry J. Cochrane review data request [pers comm]. Email to: P Eve 16 January 2012.
 Hurry J. Cochrane review data request [pers comm]. Email to: P Eve 17 January 2012.
 * Hurry J, Sylva K. Long-term outcomes of early reading intervention. *Journal of Research in Reading* 2007;**30**(3): 227–48. DOI: 10.1111/j.1467-9817.2007.00338.x
- Levy 1997** {*published data only (unpublished sought but not used)*}
 Eve P. Cochrane review data request [pers comm]. Email to: Levy B 26 September 2011.
 * Levy BA, Lysynchuk L. Beginning word recognition: benefits of training by segmentation and whole word methods. *Scientific Studies of Reading* 1997;**1**(4):359–87. DOI: 10.1207/s1532799xssr0104_4
- Levy 1999** {*published data only (unpublished sought but not used)*}
 Levy BA. Cochrane review data request [pers comm]. Email to: P Eve 4 October 2011.
 * Levy BA, Bourassa DC, Horn C. Fast and slow namers: benefits of segmentation and whole word training. *Journal of Experimental Child Psychology* 1999;**73**(2):115–38. DOI: 10.1006/jecp.1999.2497
- Lovett 1990** {*published data only (unpublished sought but not used)*}
 Frijters J. Cochrane review data request [pers comm]. Email to: P Eve 4 October 2011.
 * Lovett M, Warren-Chaplin PM, Ransby MJ, Borden SL. Training the word recognition skills of reading disabled children: treatment and transfer effect. *Journal of Educational Psychology* 1990;**82**(4):769–80. DOI: 10.1037/0022-0663.82.4.769
- Lovett 2000** {*published data only (unpublished sought but not used)*}
 Eve P. Cochrane review data request [pers comm]. Email to: J Frijters 4 October 2011.
 Lovett MW, Borden SL, DeLuca T, Lacerenza L, Benson NJ, Brackstone D. Treating the core deficits of developmental dyslexia: evidence of transfer of learning after phonologically- and strategy-based reading training programs. *Developmental Psychology* 1994;**30**(6):805–22. DOI: 10.1037/0012-1649.30.6.805
 Lovett MW, Lacerenza L, Borden SL, Frijters JC, Steinbach KA, De Palma M. Components of effective remediation for developmental reading disabilities: combining phonological and strategy-based instruction to improve outcomes. *Journal of Educational Psychology* 2000;**92**(2):263–83. DOI: 10.1037/0022-0663.92.2.263
 Lovett MW, Steinbach KA. The effectiveness of remedial programs for reading disabled children of different ages: does the benefit decrease for older children?. *Learning Disability Quarterly* 1997;**20**(3):189–210. DOI: 10.2307/1511308
 * Lovett MW, Steinbach KA, Frijters JC. Remediating the core deficits of developmental reading disability: a double-deficit perspective. *Journal of Learning Disabilities* 2000;**33**(4):334–58. DOI: 10.1177/0022219400033004006; PUBMED: 15493096
- McArthur 2015a** {*published data only*}
 McArthur G (Macquarie University, Sydney, Australia). [pers comm]. Conversation with: HC Wang (Cochrane Review Team, Sydney, Australia) 15 July 2017.
 * McArthur G, Castles A, Kohnen S, Larsen L, Jones K, Anandakumar T, et al. Sight word and phonics training

in children with dyslexia. *Journal of Learning Disabilities* 2015;**48**(4):391–407. DOI: 10.1177/0022219413504996; PUBMED: 24085229

McArthur 2015b {published data only}

McArthur G (Macquarie University, Sydney, Australia). [pers comm]. Conversation with: HC Wang (Cochrane Review Team, Sydney, Australia) 15 July 2017.
* McArthur G, Kohonen S, Jones K, Eve P, Banales E, Larsen L, et al. Replicability of sight word training and phonics training in poor readers: a randomised controlled trial. *PeerJ* 2015;**3**:e922. DOI: 10.7717/peerj.922; PMC4435451; PUBMED: 26019992

Savage 2003 {published data only (unpublished sought but not used)}

Eve P. Cochrane review data request [pers comm]. Email to: R Savage 11 February 2012.
Eve P. Cochrane review data request [pers comm]. Email to: R Savage 24 January 2012.
Savage R. Cochrane review data request [pers comm]. Email to: P Eve 30 November 2011.
* Savage R, Carless S, Stuart M. The effects of rime- and phoneme-based teaching delivered by learning support assistants. *Journal of Research in Reading* 2003;**26**(3): 211–33. DOI: 10.1111/1467-9817.00199

Savage 2005 {published data only (unpublished sought but not used)}

Eve P. Cochrane review data request [pers comm]. Email to: R Savage 11 February 2012.
Eve P. Cochrane review data request [pers comm]. Email to: R Savage 24 January 2012.
Savage R. Cochrane review data request [pers comm]. Email to: P Eve 11 February 2012.
Savage R. Cochrane review data request [pers comm]. Email to: P Eve 30 November 2011.
* Savage R, Carless S. Learning support assistants can deliver effective reading interventions for 'at-risk' children. *Educational Research* 2005;**47**(1):45–61. DOI: 10.1080/0013188042000337550

References to studies excluded from this review

About 2018 {published data only}

About KS, Barquero LA, Cutting LE. Prefrontal mediation of the reading network predicts intervention response in dyslexia. *Cortex* 2018;**101**:96–106. DOI: 10.1016/j.cortex.2018.01.009; PMC5869156; PUBMED: 29459284

Alexander 1991 {published data only}

Alexander AW, Andersen HG, Heilman PC, Voeller KK, Torgesen JK. Phonological awareness training and remediation of analytic decoding deficits in a group of severe dyslexics. *Annals of Dyslexia* 1991;**41**(1):193–206. DOI: 10.1007/BF02648086; PUBMED: 24233765

Arnold 2016 {published data only}

Arnold SS, Barton B, McArthur G, North KN, Payne JM. Phonics training improves reading in children with neurofibromatosis type 1: a prospective intervention trial. *Journal of Pediatrics* 2016;**177**:219–26.e2. DOI: 10.1016/j.jpeds.2016.06.037; PUBMED: 27480199

Berninger 2013 {published data only}

Berninger VW, Lee YL, Abbott RD, Breznitz Z. Teaching children with dyslexia to spell in a reading-writers' workshop. *Annals of Dyslexia* 2013;**63**(1):1–24. DOI: 10.1007/s11881-011-0054-0; PUBMED: 21845501

Bhide 2013 {published data only}

Bhide A, Power A, Goswami U. A rhythmic musical intervention for poor readers: a comparison of efficacy with a letter-based intervention. *Mind, Brain, and Education* 2013;**7**(2):113–23. DOI: 10.1111/mbe.12016

Christodoulou 2017 {published data only}

Christodoulou JA, Cyr A, Murtagh J, Chang P, Lin J, Guarino AJ, et al. Impact of intensive summer reading intervention for children with reading disabilities and difficulties in early elementary school. *Journal of Learning Disabilities* 2017;**50**(2):115–27. DOI: 10.1177/0022219415617163; PUBMED: 26712799

Dubois 2014 {published data only}

Dubois MR, Volpe RJ, Hemphill EM. A randomized trial of a computer-assisted tutoring program targeting letter-sound expression. *School Psychology Review* 2014;**43**(2): 210–21. ERIC Number: EJ1142183]

Foorman 1997 {published data only}

Foorman BR, Francis DJ, Winikates D, Mehta P, Schatschneider C, Fletcher JM. Early interventions for children with reading disabilities. *Scientific Studies of Reading* 1997;**1**(3):255–76. DOI: 10.1207/s1532799xssr0103_5

Foorman 1998 {published data only}

Foorman BR, Francis DJ, Fletcher JM, Schatschneider C, Mehta P. The role of instruction in learning to read: preventing reading failure in at-risk children. *Journal of Educational Psychology* 1998;**90**(1):37–55. DOI: 10.1037/0022-0663.90.1.37

Gillon 1997 {published data only}

Gillon G, Dodd B. Enhancing the phonological processing skills of children with specific reading disability. *European Journal of Disorders of Communication* 1997;**32**(2):67–90. PUBMED: 9279428]

Gillon 2000 {published data only}

Gillon GT. The efficacy of phonological awareness intervention for children with spoken language impairment. *Language, Speech, and Hearing Services in Schools* 2000;**31**(2):126–41. DOI: 10.1044/0161-1461.3102.126; PUBMED: 27764385

Gillon 2002 {published data only}

Gillon GT. Follow-up study investigating with benefits of phonological awareness intervention for children with spoken language impairment. *International Journal of Language & Communication Disorders* 2002;**37**(4):381–400. DOI: 10.1080/1368282021000007776; PUBMED: 12396840

Goldstein 2017 {published data only}

Goldstein H, Olszewski A, Haring C, Greenwood CR, McCune L, Carta J, et al. Efficacy of a supplemental phonemic awareness curriculum to instruct preschoolers with delays in early literacy development. *Journal of Speech,*

- Language, and Hearing Research* 2017;**60**(1):89–103. DOI: 10.1044/2016.JSLHR-L-15-0451
- Gorard 2015** *{published data only}*
Gorard S, Siddiqui N, See BH. Fresh start: evaluation report and executive summary. v1.educationendowmentfoundation.org.uk/uploads/pdf/Fresh_Start_(Final).pdf (accessed prior to 11 September 2018).
- Hatcher 1994** *{published data only}*
Hatcher PJ, Hulme C, Ellis AW. Ameliorating early reading failure by integrating the teaching of reading and phonological skills: the phonological linkage hypothesis. *Child Development* 1994;**65**(1):41–57. DOI: 10.1111/j.1467-8624.1994.tb00733.x
- Hatcher 2006** *{published data only}*
Hatcher PJ, Goetz K, Snowling MJ, Hulme C, Gibbs S, Smith G. Evidence for the effectiveness of the Early Literacy Support programme. *British Journal of Educational Psychology* 2006;**76**(Pt 2):351–67. DOI: 10.1348/000709905X39170; PUBMED: 16719968
- Jeffes 2016** *{published data only}*
Jeffes B. Raising the reading skills of secondary-age students with severe and persistent reading difficulties: evaluation of the efficacy and implementation of a phonics-based intervention programme. *Educational Psychology in Practice* 2016;**32**(1):73–84. DOI: 10.1080/02667363.2015.1111198
- King 2015** *{published data only}*
King B, Kasim A. Rapid phonics: evaluation report and executive summary. v1.educationendowmentfoundation.org.uk/uploads/pdf/Rapid_Phonics_(Final).pdf (accessed prior to 11 September 2018).
- Lovett 1988** *{published data only}*
Lovett MW, Ransby MJ, Barron RW. Treatment, subtype, and word type effects in dyslexic children's response to remediation. *Brain and Language* 1988;**34**(2):328–49. PUBMED: 3401697]
- Lovett 1989** *{published data only}*
Lovett MW, Ransby MJ, Hardwick N, Johns MS, Donaldson SA. Can dyslexia be treated? Treatment-specific and generalized treatment effects in dyslexic children's response to remediation. *Brain and Language* 1989;**37**(1): 90–121. PUBMED: 2752277]
- Lovett 1994** *{published data only}*
Lovett MW, Barron RW, Forbes JE, Cuksts B, Steinbach KA. Computer speech-based training of literacy skills in neurologically impaired children: a controlled evaluation. *Brain and Language* 1994;**47**(1):117–54. DOI: 10.1006/brln.1994.1045; PUBMED: 7922474
- Lovett 2012** *{published data only}*
Lovett MW, Lacerenza L, De Palma M, Frijters JC. Evaluating the efficacy of remediation for struggling readers in high school. *Journal of Learning Disabilities* 2012;**45**(2): 151–69. DOI: 10.1177/0022219410371678; PUBMED: 22183192
- Merrell 2015** *{published data only}*
Merrell C, Kasim A. Butterfly phonics: evaluation report and executive summary. files.eric.ed.gov/fulltext/ED581118.pdf (accessed prior to 11 September 2018).
- Messer 2018** *{published data only}*
Messer D, Nash G. An evaluation of the effectiveness of a computer-assisted reading intervention. *Journal of Research in Reading* 2018;**41**(1):140–58. DOI: 10.1111/1467-9817.12107
- Metsala 2017** *{published data only}*
Metsala JL, David MD, Brown S. An examination of reading skills and reading outcomes for youth involved in a crime prevention program. *Reading & Writing Quarterly* 2017;**33**(6):549–62. DOI: 10.1080/10573569.2016.1268081
- Munro 2017** *{published data only}*
Munro J. Who benefits from which reading intervention in the primary years? Match the intervention with the reading profile. *Australian Journal of Learning Difficulties* 2017;**22**(2):133–51. DOI: 10.1080/19404158.2017.1379027
- Olson 1992** *{published data only}*
Olson RK, Wise BW. Reading on the computer with orthographic and speech feedback: an overview of the Colorado Remediation Project. *Reading and Writing* 1992;**4**(2):107–44. DOI: 10.1007/BF01027488
- Olson 1997** *{published data only}*
Olson RK, Wise B, Ring J, Johnson M. Computer-based remedial training in phoneme awareness and phonological decoding: effects on the posttraining development of word recognition. *Scientific Studies of Reading* 1997;**1**(3):235–53. DOI: 10.1207/s1532799xssr0103_4
- Rashotte 2001** *{published data only}*
Rashotte CA, MacPhee K, Torgesen JK. The effectiveness of a group reading instruction program with poor readers in multiple grades. *Learning Disability Quarterly* 2001;**24**(2): 119–34. DOI: 10.2307/1511068
- Savage 2018** *{published data only}*
Savage R, Georgiou G, Parrila R, Maiorino K. Preventative reading interventions teaching direct mapping of graphemes in texts and set-for-variability aid at-risk learners. *Scientific Studies of Reading* 2018;**22**(3):225–47. DOI: 10.1080/10888438.2018.1427753
- Schaars 2017** *{published data only}*
Schaars MM, Segers E, Verhoeven L. Word decoding development during phonics instruction in children at risk for dyslexia. *Dyslexia* 2017;**23**(2):141–60. DOI: 10.1002/dys.1556; PMC6084288; PUBMED: 28470910
- Schlesinger 2017** *{published data only}*
Schlesinger NW, Gray S. The impact of multisensory instruction on learning letter names and sounds, word reading, and spelling. *Annals of Dyslexia* 2017;**67**(3): 219–58. DOI: 10.1007/s11881-017-0140-z; PUBMED: 28255950
- Seiler 2018** *{published data only}*
Seiler A, Leitão S, Blöschfelds M. WordDriver-1: evaluating the efficacy of an app-supported decoding intervention for children with reading impairment. *International Journal*

- of Language & Communication Disorders 2018 Apr 24 [Epub ahead of print]. DOI: 10.1111/1460-6984.12388; PUBMED: 29691983
- Steady 2016** *{published data only}*
Steady LM, Elleman AM, Lovett MW, Compton DL. Exploring differential effects across two decoding treatments on item-level transfer in children with significant word reading difficulties: a new approach for testing intervention elements. *Scientific Studies of Reading* 2016;**20**(4):283–95. DOI: 10.1080/10888438.2016.1178267; NIHMSID: NIHMS804588; PMC5460658; PUBMED: 28596701
- Storey 2017** *{published data only}*
Storey C, McDowell C, Leslie JC. Evaluating the efficacy of the Headsprout® reading program with children who have spent time in care. *Behavioral Interventions* 2017;**32**(3): 285–93. DOI: 10.1002/bin.1476
- Torgesen 1997** *{published data only}*
Torgesen JK, Wagner RK, Rashotte CA. Prevention and remediation of severe reading disabilities: keeping the end in mind. *Scientific Studies of Reading* 1997;**1**(3):217–34. DOI: 10.1207/s1532799xssr0103_3
- Torgesen 1999a** *{published data only}*
Torgesen JK, Wagner RK, Rashotte CA, Rose E, Lindamood P, Conway T, et al. Preventing reading failure in young children with phonological processing disabilities: group and individual responses to instruction. *Journal of Educational Psychology* 1999;**91**(4):579–93. DOI: 10.1037/0022-0663.91.4.579
- Torgesen 2001** *{published data only}*
Torgesen JK, Alexander AW, Wagner RK, Rashotte CA, Voeller KK, Conway T. Intensive remedial instruction for children with severe reading disabilities: immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities* 2001;**34**(1):33–58, 78. DOI: 10.1177/002221940103400104; PUBMED: 15497271
- Torgesen 2006** *{published data only}*
Torgesen J, Myers D, Schirm A, Stuart E, Vartivarian S, Mansfield W, et al. Closing the reading gap: first year findings from a randomized trial of four reading interventions for striving readers. Final report. Washington (DC): Corporation for the Advancement of Policy Evaluation; 2006 February. Reference No.: 8970–400.
- Van Gorp 2017** *{published data only}*
Van Gorp K, Segers E, Verhoeven L. The role of feedback and differences between good and poor decoders in a repeated word reading paradigm in first grade. *Annals of Dyslexia* 2017;**67**(1):1–25. DOI: 10.1007/s11881-016-0129-z; PMC5346118; PUBMED: 27068186
- Vellutino 1986** *{published data only}*
Vellutino FR, Scanlon DM. Experimental evidence for the effects of instructional bias on word identification. *Exceptional Children* 1986;**53**(2):145–55. PUBMED: 3770059]
- Vellutino 1987** *{published data only}*
Vellutino FR, Scanlon DM. Phonological coding, phonological awareness, and reading ability: evidence from a longitudinal and experimental study. *Merrill-Palmer Quarterly* 1987;**33**(3):321–63. EJ361477]
- Vellutino 1996** *{published data only}*
Vellutino FR, Scanlon DM, Sipay ER, Small SG, Pratt A, Chen R, et al. Cognitive profiles of difficult-to-remediate and readily remediated poor readers: early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology* 1996;**88**(4):601–38. DOI: 10.1037/0022-0663.88.4.601
- Wheldall 2017** *{published data only}*
Wheldall K, Wheldall R, Madeline A, Reynolds M, Arakelian S. Further evidence for the efficacy of an evidence-based, small group, literacy intervention program for young struggling readers. *Australian Journal of Learning Difficulties* 2017;**22**(1):3–13. DOI: 10.1080/19404158.2017.1287102
- Wise 1995** *{published data only}*
Wise BW, Olson RK. Computer-based phonological awareness and reading-instruction. *Annals of Dyslexia* 1995; **45**(1):97–122. DOI: 10.1007/BF02648214; PUBMED: 24234190
- Wise 1997** *{published data only}*
Wise B, Ring J, Sessions L, Olson RK. Phonological awareness with and without articulation: a preliminary study. *Learning Disability Quarterly* 1997;**20**(3):211–25. DOI: 10.2307/1511309
- Wise 1999** *{published data only}*
Wise BW, Ring J, Olson RK. Training phonological awareness with and without explicit attention to articulation. *Journal of Experimental Child Psychology* 1999;**72**(4): 271–304. DOI: 10.1006/jecp.1999.2490; PUBMED: 10074381
- Wise 2000** *{published data only}*
Wise BW, Ring J, Olson RK. Individual differences in gains from computer-assisted remedial reading. *Journal of Experimental Child Psychology* 2000;**77**(3):197–235. DOI: 10.1006/jecp.1999.2559; PUBMED: 11023657

Additional references

Alexander-Passe 2015

Alexander-Passe N. Dyslexia: investigating self-harm and suicidal thoughts/attempts as a coping strategy. *Journal of Psychology & Psychotherapy* 2015;**5**(6):1. DOI: 10.4172/2161-0487.1000224

Blachman 2000

Blachman BA. Phonological awareness. In: Kamil ML, Mosenthal PB, Pearson PD, Barr R editor(s). *Handbook of Reading Research*. Vol. **III**, Mahwah (NJ): Lawrence Erlbaum, 2000:483–502.

Bus 1999

Bus A, Van IJzendoorn MH. Phonological awareness and early reading: a meta-analysis of experimental training studies. *Journal of Educational Psychology* 1999;**91**(3): 403–14. DOI: 10.1037/0022-0663.91.3.403

- Carroll 2006**
Carroll JM, Iles JE. An assessment of anxiety levels in dyslexic students in higher education. *British Journal of Educational Psychology* 2006;**76**(Pt 3):651–62. DOI: 10.1348/000709905X66233; PUBMED: 16953967
- Castles 1993**
Castles A, Coltheart M. Varieties of developmental dyslexia. *Cognition* 1993;**47**(2):149–80. [PUBMED: 8324999]
- Chall 1967**
Chall JS. *Learning to Read: The Great Debate*. New York (NY): McGraw-Hill, 1967.
- Chen 2016**
Chen V, Savage RS. Corrigendum to the 'Evidence for a simplicity principle: teaching common complex grapheme-to-phonemes improves reading and motivation in at-risk readers' *Journal of Research in Reading*, 2014;**37**(2): 196–214. *Journal of Research in Reading* 2016;**39**(1):126–31. DOI: 10.1111/1467-9817.12068
- Clay 1985**
Clay MM. *The Early Detection of Reading Difficulties: A Diagnostic Survey with Recovery Procedures*. 3. Auckland (NZ): Heinemann, 1985.
- Cohen 1988**
Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd Edition. Hillsdale (NJ): Lawrence Erlbaum Associates, 1988.
- Coltheart 2001**
Coltheart M, Rastle K, Perry C, Langdon R, Ziegler J. DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological Review* 2001;**108**(1):204–56. [PUBMED: 11212628]
- Daniel 2006**
Daniel SS, Walsh AK, Goldston DB, Arnold EM, Reboussin BA, Wood FB. Suicidality, school dropout, and reading problems among adolescents. *Journal of Learning Disabilities* 2006;**39**(6):507–14. DOI: 10.1177/00222194060390060301; PUBMED: 17165618
- Dawson 2000**
Dawson L, Venn ML, Gunter PL. The effects of teacher versus computer reading models. *Behavioral Disorders* 2000;**25**(2):105–13. [EJ603392]
- Deeks 2011**
Deeks JJ, Higgins JPT, Altman DG. Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JP, Green S, editor(s). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from handbook.cochrane.org.
- Ehri 2001**
Ehri LC, Nunes SR, Stahl SA, Willows DM. Systematic phonics instruction helps students learn to read: evidence from the national reading panel's meta-analysis. *Review of Educational Research* 2001;**71**(3):393–447. DOI: 10.3102/00346543071003393
- Ehri 2014**
Ehri LC. Orthographic mapping in the acquisition of sight word reading, spelling memory, and vocabulary learning. *Scientific Studies of Reading* 2014;**18**(1):5–21. DOI: 10.1080/10888438.2013.819356
- Elbaum 2000**
Elbaum B, Vaughn S, Hughs MT, Moody SW. How effective are one-on-one tutoring programs in reading for elementary students at risk for reading failure? A meta-analysis of the intervention research. *Journal of Educational Research* 2000;**92**(4):605–19. [EJ621002]
- Elliot 1983**
Elliot CD. *British Ability Scales*. Windsor (UK): NFER-Nelson, 1983.
- Fletcher 2005**
Fletcher JM, Denton C, Francis DJ. Validity of alternative approaches for the identification of learning disabilities: operationalizing unexpected underachievement. *Journal of Learning Disabilities* 2005;**38**(6):545–52. DOI: 10.1177/00222194050380061101; 16392697
- Galushka 2014**
Galushka K, Ise E, Krick K, Schulte-Körne G. Effectiveness of treatment approaches for children and adolescents with reading disabilities: a meta-analysis of randomised controlled trials. *PLoS One* 2014;**9**(2):e89900. DOI: 10.1371/journal.pone.0089900; 24587110; PMC3935956
- Goldman 1974**
Goldman R, Fristoe M, Woodcock RW. *G-F-W Diagnostic Auditory Discrimination Test*. Circle Pines (MN): American Guidance Service, 1974.
- Harm 1999**
Harm M, Seidenberg M. Reading acquisition, phonology, and dyslexia: insights from a connectionist model. *Psychological Review* 1999;**106**:491–528.
- Higgins 2011**
Higgins JP, Altman DG, Sterne JA. Chapter 8: Assessing risk of bias in included studies. In: Higgins JP, Green S, editor(s). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from handbook.cochrane.org.
- Hurford 1993**
Hurford DP, Darrow LJ, Edwards TL, Howerton CJ, Mote CR, Schauf JD, et al. An examination of phonemic processing abilities in children during their first-grade year. *Journal of Learning Disabilities* 1993;**26**(3):167–77. DOI: 10.1177/002221949302600304
- Karolson 2009**
Karolson CW, Rapoff MA. Attrition in randomized controlled trials for pediatric chronic conditions. *Journal of Pediatric Psychology* 2009;**34**(7):782–93. DOI: 10.1093/jpepsy/jsn122; PUBMED: 19064607
- Lefebvre 2011**
Lefebvre C, Manheinmer E, Glanville J. Chapter 6: Searching for studies. In: Higgins JP, Green S, editor

- (s). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from handbook.cochrane.org.
- MacGinitie 2002**
MacGinitie WH, MacGinitie RK, Maria K, Dreyer LG, Hughes KE. *Gates MacGinitie Reading Tests, Forms S & T*. 4th Edition. Boston (MA): Houghton Mifflin, 2002.
- Maughan 2003**
Maughan B, Rowe R, Loeber R, Stouthamer-Loeber M. Reading problems and depressed mood. *Journal of Abnormal Child Psychology* 2003;**31**(2):219–29. [PUBMED: 12735404]
- McArthur 2013**
McArthur GM, Jones K, Anandakumar T, Larsen L, Castles A, Coltheart M. A Test of Everyday Reading Comprehension (TERC). *Australian Journal of Learning Difficulties* 2013;**18**(1):35–85. DOI: 10.1080/19404158.2013.779588
- McArthur 2016**
McArthur G, Castles A, Kohonen S, Banales E. Low self-concept in poor readers: prevalence, heterogeneity, and risk. *PeerJ* 2016;**4**:e2669. DOI: 10.7717/peerj.2669; PMC5111895; PUBMED: 27867764
- Neale 1988**
Neale MD. *Neale Analysis of Reading Ability, Revised*. 3rd Edition. Hawthorn (VI): ACER Press, 1988.
- Perry 2007**
Perry C, Ziegler JC, Zorzi M. Nested incremental modeling in the development of computational theories: the CDP+ model of reading aloud. *Psychological Review* 2007;**114**(2):273–315. DOI: 10.1037/0033-295X.114.2.273; PUBMED: 17500628
- Review Manager 2014 [Computer program]**
Nordic Cochrane Centre, The Cochrane Collaboration. Review Manager (RevMan). Version 5.3. Copenhagen: Nordic Cochrane Centre, The Cochrane Collaboration, 2014.
- Ryan 2016**
Ryan R, Santesso N, Hill S. Preparing summary of findings (SoF) tables. figshare.com/articles/Summary_of_findings_tables/6818891 (accessed prior to 11 September 2018).
- Saghaei 2011**
Saghaei M, Saghaei S. Implementation of an open-source customizable minimization program for allocation of patients to parallel groups in clinical trials. *Journal of Biomedical Science and Engineering* 2011;**4**:734–9. DOI: 10.4236/jbise.2011.411090
- Schünemann 2011a**
Schünemann HJ, Oxman AD, Higgins JP, Vist GE, Glasziou P, Guyatt GH. Chapter 11: Presenting results and 'Summary of findings' tables. In: Higgins JP, Green S, editor(s). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from handbook.cochrane.org.
- Schünemann 2011b**
Schünemann HJ, Oxman AD, Vist GE, Higgins JP, Deeks JJ, Glasziou P, et al. Chapter 12: Interpreting results and drawing conclusions. In: Higgins JP, Green S, editor(s). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from handbook.cochrane.org.
- Share 1995**
Share D. Phonological recoding and self-teaching: sine qua non of reading acquisition. *Cognition* 1995;**55**(2):151–218; 219–26. [PUBMED: 7789090]
- Shaywitz 1992**
Shaywitz SE, Escobar MD, Shaywitz BA, Fletcher JM, Makuch R. Evidence that dyslexia may represent the lower tail of a normal distribution of reading ability. *New England Journal of Medicine* 1992;**326**(3):145–50. DOI: 10.1056/NEJM199201163260301; PUBMED: 1727544
- Shaywitz 2001**
Shaywitz SE, Shaywitz BA. The neurobiology of reading and dyslexia. *Focus on Basics* 2001;**5**(A):11–5.
- Shultz 2010**
Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *Annals of Internal Medicine* 2010;**152**(11):726–32. DOI: 10.7326/0003-4819-152-11-201006010-00232
- Stahl 1994**
Stahl SA, Murray BA. Defining phonological awareness and its relationship to early reading. *Journal of Educational Psychology* 1994;**86**(2):221–34. DOI: 10.1037/0022-0663.86.2.221
- Sterne 2011**
Sterne JA, Egger M, Moher D. Chapter 10: Addressing reporting biases. In: Higgins JP, Green S, editor (s). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from handbook.cochrane.org.
- Suggate 2010**
Suggate SP. Why what we teach depends on when: grade and reading intervention modality moderate effect size. *Developmental Psychology* 2010;**46**(6):1556–79. DOI: 10.1037/a0020612; PUBMED: 20873927
- Swanson 1999**
Swanson HL, Hoskyn M, Lee C. *Interventions for Students with Learning Disabilities*. New York (NY): Guilford Press, 1999.
- Therrien 2004**
Therrien WJ. Fluency and comprehension gains as a result of repeated reading: a meta-analysis. *Remedial and Special Education* 2004;**25**(4):252–61. [EJ695617]
- Torgesen 1999b**
Torgesen JK, Rashotte CA, Wagner RK. *Test of Word Reading Efficiency (TOWRE)*. Austin (TX): Pro-Ed, 1999.

Torgesen 2010

Torgesen JK, Wagner RK, Rachotte CA, Herron J, Lindamood P. Computer-assisted instruction to prevent early reading difficulties in students at risk for dyslexia: outcomes from two instructional approaches. *Annals of Dyslexia* 2010;**60**(1):40–56. DOI: 10.1007/s11881-009-0032-y; PMC2888606 ; PUBMED: 20052566

Vousden 2008

Vousden JI. Units of English spelling-to-sound mapping: a rational approach to reading instruction. *Applied Cognitive Psychology* 2008;**22**(2):247–72. DOI: 10.1002/acp.1371

Wechsler 2001

Wechsler D. *Wechsler Individual Achievement Test*. San Antonio (TX): Psychological Corporation, 2001.

Weekes 1997

Weekes BS. Differential effects of number of letters on word and nonword naming latency. *Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology* 1997;**50**(2):439–56. DOI: 10.1080/713755710

Williams 2010

Williams KT. *Group Reading Assessment and Diagnostic Evaluation: Level 1, Version A*. Melbourne (VI): Pearson,

2010.

Woodcock 1987

Woodcock RW. *Woodcock Reading Mastery Tests - Revised*. Circle Pines (MN): American Guidance Service, 1987.

Woodcock 2001

Woodcock RW, McGrew KS, Mather N. *Woodcock-Johnson III*. 3rd Edition. Itasca (IL): Riverside Publishing, 2001.

References to other published versions of this review**McArthur 2011**

McArthur G, Castles A, Kohnen S, Larsen L, Jones K, Anandakumar T, et al. Phonics training for English-speaking poor readers. *Cochrane Database of Systematic Reviews* 2011, Issue 5. DOI: 10.1002/14651858.CD009115

McArthur 2012

McArthur G, Eve PM, Jones K, Banales E, Kohnen S, Anandakumar T, et al. Phonics training for English-speaking poor readers. *Cochrane Database of Systematic Reviews* 2012, Issue 12. DOI: 10.1002/14651858.CD009115.pub2; PUBMED: 23235670

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

Characteristics of included studies [ordered by study ID]

Barker 1995

Methods	Randomised controlled trial 2 intervention groups (phonics, phonological awareness (not relevant)) and 1 control group (alternative training)
Participants	<p>Location/setting: 2 elementary schools; USA</p> <p>Criteria: score \leq 40th percentile on the WJRT Word Identification subtest; score $<$ 50th percentile on the Sound Categorisation subtest</p> <p>Recruits: 54 English-speaking children, who scored slightly below mean range on Vocabulary subtest from Stanford Binet IV-Revised (mean 16.5, SD 2.36; range 11-22)</p> <p>Sex: not reported</p> <p>Mean age: not reported (SD not reported; range 6 years, 2 months to 7 years, 8 months)</p> <p>Ethnicity: not reported</p> <p>Sample size: 32 English-speaking children</p> <p>Allocation: “Children were randomly assigned to one of three conditions” (quote, p 95). This review used the phonological decoding training group as the intervention group and the maths training group as the control group. There was also a phonological awareness control group (see notes), which was not used by this review</p> <p>Intervention groups:</p> <ol style="list-style-type: none"> 1. phonics: n = 18 (mean age, SD, and range not reported) 2. phonological awareness: n = 18 (mean age, SD, and range not reported) <p>Control group: n = 18 (mean age, SD, and range not reported)</p>
Interventions	<p>Intervention:</p> <ol style="list-style-type: none"> 1. phonics training: phonological decoding training: Hint and Hunt I programme: “designed to acquaint children with the basic short vowel sounds and provide practice in identifying words containing those sounds” (quote, p 94; phonics) 2. phonological awareness: phoneme awareness training using Daisy Quest <p>Control: attentional control group: maths-oriented software programs (Alien Addition, Math Rabbit, Math Blaster)</p> <p>Procedure: training took place in school psychologist’s office. Groups of 3 and 4 throughout the school day. 25-minute sessions, 4 times/week (Monday to Thursday) for 8 weeks. Friday used as make-up sessions. 1 experimenter at each site who set up each station with appropriate programme for each student. Training done via computer. Experimenter helped with technical issues but no conceptual issues. Students rewarded with 1 sticker at end of session</p>
Outcomes	<p>Time of post-test: immediately after training completed</p> <p>Primary and secondary outcomes: non-word reading accuracy (Word Analysis subtest from WJRT), regular and irregular word reading accuracy (Word Identification subtest from WJRT), and phonological awareness (experimental: phoneme elision)</p>
Notes	<ol style="list-style-type: none"> 1. The phonological awareness training group used <i>Daisy Quest</i> and <i>Daisy’s Castle</i>. <i>Daisy Quest</i> trains recognising words that rhyme; recognising words that have the same beginning, middle, and ending sounds. <i>Daisy’s Castle</i> teaches these additional skills:

Barker 1995 (Continued)

recognising words formed from a series of phonemes presented as onset and rime; recognising words that can be formed from a series of separately presented phonemes; counting the number of sounds in words. These programmes did not include phonics and so were not included as an intervention in this review.

2. 2 measures were used to test reading accuracy: non-words. We only included the Word Analysis subtest of the WJRT as it is a published test with known reliability.

3. 2 measures were used to test reading accuracy: words. We just used Word Identification of the WJRT to represent mixed/regular word reading accuracy.

Study start and end dates: not reported
Funding: not reported
Potential/declared conflicts of interest: none reported

<i>Risk of bias</i>		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Quote from publication: "Children were randomly assigned to one of three conditions" (p 95) Comment: no other information provided
Allocation concealment (selection bias)	Low risk	Comment: could not foresee assignment due to central allocation of participants to groups
Blinding of participants and personnel (performance bias) All outcomes	Unclear risk	Comment: no information provided; however, participants were children with little understanding of reading treatment techniques and hence were unlikely to understand allocation
Blinding of outcome assessment (detection bias) All outcomes	Unclear risk	Comment: no information provided; however, this study used objective tests of literacy-related skills that are designed to avoid assessor bias
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Comment: no information provided; allocated group sizes not reported in publication, and no response to request for information
Selective reporting (reporting bias)	Unclear risk	Comment: data reported for all phonological and reading tests listed in methods; adequate detail for data to be included in analysis
Other bias	Low risk	Comment: none apparent

Methods	Randomised controlled trial 1 intervention group (phonics + phonological awareness) and 1 control group (untrained)
Participants	<p>Location/setting: medium-sized private primary school in Western Sydney, Australia</p> <p>Criteria: received weekly group-based remedial reading instruction at the school and referred to the study by a support teacher</p> <p>Recruits: the participants had no other comorbid specific learning disorders. They had a mean delay of 13 months on a word reading task (subtest on WIAT-II); 11 months on a reading comprehension task (subtest on WIAT-II), and 25 months on a pseudoword decoding task (subtest on WIAT-II). Participants had a mean FSIQ of 100.15 (SD 9.38)</p> <p>Sex: 15 males; 5 females</p> <p>Mean age: 101.35 months (SD 17.58 months; range not reported)</p> <p>Ethnicity: not reported</p> <p>Sample size: 20 English-speaking dyslexic primary students</p> <p>Allocation: random allocation</p> <p>Intervention group: n = 10 (mean age 99.8 months; SD 18.94; range not reported)</p> <p>Control group: n = 10 (mean age 102.9 months; SD 16.98; range not reported)</p>
Interventions	<p>Intervention: <i>Phonics Alive! 2: The Sound Blender</i> (version 1.2): 10-week training programme. “Program consists of 12 modules which systematically build skills in phoneme awareness, phoneme-grapheme correspondences, sound and letter blending, and speed of processing” (quote, p 41)</p> <p>Control: students continued to receive their school-based reading instruction (both in-class and at a weekly remedial group with the support teacher)</p> <p>Procedure: children in the intervention group continued their school-based instruction while they did their training at home and at school on a computer. At home, each training module took approximately 15 minutes to complete. Students were instructed to repeat each module until they reached a mastery level of 90% correct. Upon mastery of a module, students had to complete review worksheets. According to parents, a mean of 3.6 computer modules were attempted per child per week. “Thus, over the 10-week training period, students completed a mean of 46 module attempts which represented approximately 11.5 hours of on-computer time” (quote) (in addition to 30 minutes/week with researcher: 5 hours). At school, children did “a weekly, 30 minute, one-on-one session with the researcher where the student’s progress was assessed by reviewing their progress chart and completed worksheets (5 minutes) and completing the current module on a computer (to verify mastery).” (quote) Any remaining time was spent playing a “nonsense word game” (quote) (p 41)</p>
Outcomes	<p>Time of post-test: immediately after training completed</p> <p>Primary outcomes: non-word reading accuracy (WIAT-II: Pseudoword Decoding subtest), regular and irregular word reading accuracy (WIAT-II: Word Reading subtest) and reading comprehension (WIAT-II subtest)</p>
Notes	<p>Contacted author for post-test SDs</p> <p>Study start and end dates: not reported</p> <p>Funding: not reported</p> <p>Potential/declared conflicts of interest: none reported</p>

<i>Risk of bias</i>		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	<p>Quote from publication: “participants were randomly assigned to either a control or treatment condition” (p 41)</p> <p>Quote from personal communication: “participants who met selection criteria were randomly assigned to either the Tx [treatment] or Ct [control] condition by drawing eligible names from a hat and placing sequentially into Tx/Ct until all were assigned.”</p>
Allocation concealment (selection bias)	Low risk	<p>Comment: could not foresee assignment due to central allocation of participants to groups</p>
Blinding of participants and personnel (performance bias) All outcomes	Low risk	<p>Quote from personal communication: “given this was a simple Tx/Ct [treatment/control] design there was no way to blind study participants or personnel from knowledge of who was in the treatment group. However, the trainer had no previous knowledge or awareness of the participants and was not involved in the referral process (they were referred by the school counsellor).”</p> <p>Comment: participants were children with little understanding of reading treatment techniques and hence were unlikely to understand allocation</p>
Blinding of outcome assessment (detection bias) All outcomes	Unclear risk	<p>Quote from personal communication: “Initial assessment of IQ and reading was conducted by the investigator on all participants PRIOR to their random assignment to Tx [treatment] or Ct [control] conditions, and thus the assessor was unaware of their future status in the study. Given this was a pilot study, post-treatment assessment was conducted by the same assessor on all students and this precluded the assessor conducting blind post-treatment assessments.”</p> <p>Comment: study used objective tests of literacy-related skills that are designed to avoid assessor bias</p>

Blythe 2006 (Continued)

Incomplete outcome data (attrition bias) All outcomes	Low risk	Comment: DF values indicated that data for all randomised participants were included in the analyses. Author sent post-test SD
Selective reporting (reporting bias)	Unclear risk	Comment: data reported for all reading tests listed in methods; adequate detail for data to be included in analysis
Other bias	Low risk	Comment: none apparent

Chen 2014

Methods	Randomised controlled trial 1 intervention group (phonics + sight word) and 1 control group (alternative training)
Participants	<p>Location/setting: grade 2 students from a regular public sector in Quebec, Canada</p> <p>Criteria: grade 2 students, considered to be 'at-risk readers,' who fell 1 SD below mean on the GRADE (standardised test). Grade 2 (mean 2.89; SD 0.90). Using the GRADE, a stanine score was calculated corresponding to the raw scores. Grade 2 stanine score (mean 5; SD 2)</p> <p>Recruits: 18 grade 2 students</p> <p>Sex: 7 male; 11 female</p> <p>Mean age: 7.06 years (SD 0.24; range 7-8 years)</p> <p>Ethnicity: bilingual speakers of English and French</p> <p>Sample size: 18 bilingual (French and English) students</p> <p>Allocation: stratified randomisation - participants were matched with another participant in the same class who scored similarly and in order of predetermined importance on the assessments: "phonemic blending assessment, word recognition test from the GRADE, spelling test and reading motivation" (quote, p 202). Within the pair, participants were randomly assigned to either GPC or word usage group using an online random number generator (www.random.org).</p> <p>Intervention group: n = 9 complex GPC group (mean age, SD, and range not reported)</p> <p>Control group: n = 9 word usage group (mean age, SD, and range not reported)</p>
Interventions	<p>Intervention: complex GPC and sight word training. Phonics was taught in the context of words (both regular and irregular). For example, the sound /sh/ was taught in the word /she/. Participants were shown a target GPC within a target word which was written in a different colour to the other letters, pulled out the target GPC from the word using physical letters, heard the word in text as the researcher read a story, had to identify the words in the text containing the target GPC, then had to read the target word aloud after. The researcher also explained to the participants where the GPC is usually located within words</p> <p>Control: word usage condition. Lessons focused on the usage of target words in sentences through sentence activities where participants had to use the target word in the correct way, and then by writing sentences that used the target word. Review sessions occurred 3 times, each after 10 words were taught, and then again on the final day to review all words that were taught</p>

	Procedure: 20 minutes/group (4-5 students) outside the classroom. 3-4 sessions/week for 9 consecutive weeks, with a total of 30 sessions. 600 minutes total (or 10 hours)	
Outcomes	<p>Time of post-test: not explicitly reported but likely immediately</p> <p>Primary outcomes: accuracy word reading (word recognition for words with taught GPCs; word recognition assessment from GRADE), accuracy word reading (word recognition for all words), spelling (experimental test: spell 9 words that do not contain target GPC), and blending (at pretest only; phonemic blending test by Pennington Publishing www.penningtonpublishing.com)</p> <p>Secondary outcome: reading motivation (reading and self-concept scale)</p>	
Notes	<ol style="list-style-type: none"> Grade 1 students were also reported in the study; however, they did not meet criteria for 'at-risk' readers and were therefore not included in this review. Correspondence with authors for more information on intervention: "Our aim was to use regular words so the students could transfer their learning to other words containing the same GPC" (quote). The researcher also conducted > 4 hours of informal classroom observations. In the classroom, GPC units were taught on the basis of individual words rather than as a theme, as is taught in the intervention. Chief investigator asked for further information to clarify regularity of words used in GPC training and confirmation of group size. <p>Study start and end dates: not reported</p> <p>Funding: not reported</p> <p>Potential/declared conflicts of interest: none reported</p>	
Risk of bias		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Comment: participants randomly assigned to group using online random number generator
Allocation concealment (selection bias)	Low risk	Comment: could not foresee assignment due to central allocation of participants to groups
Blinding of participants and personnel (performance bias) All outcomes	Unclear risk	Comment: no information provided; however, participants were children with little understanding of reading treatment techniques and hence were unlikely to understand allocation
Blinding of outcome assessment (detection bias) All outcomes	Unclear risk	Comment: no information provided; however, this study used objective tests of literacy-related skills that were designed to avoid assessor bias

Chen 2014 (Continued)

Incomplete outcome data (attrition bias) All outcomes	Low risk	Comment: no information provided; however, Table 2 suggested that all 38 participants contributed data at pre- and post-test, suggesting no attrition
Selective reporting (reporting bias)	Unclear risk	Comment: phonemic blending was used to select and match groups. Data reported pretest but no data available at post-test
Other bias	Low risk	Comment: none apparent

Ford 2009

Methods	Randomised controlled trial 1 intervention group (phonics + phoneme awareness) and 1 control group (untrained)
Participants	<p>Location/setting: an alternative high school in Illinois, USA</p> <p>Criteria: enrolled in a remedial reading programme</p> <p>Recruits: 20 English-speaking participants. Most participants were bilingual. Mostly Title 1 (lower SES)</p> <p>Sex: 9 male; 11 female</p> <p>Mean age: 16.18 years (SD and range not reported)</p> <p>Ethnicity: 4 African-American, 12 Hispanic, 2 white</p> <p>Sample size: 18 English-speaking participants</p> <p>Allocation: “students were randomly assigned to an experimental or control group by drawing names” (quote, p 49)</p> <p>Intervention group: n = 9 (female = 5, male = 4); mean age 16 years, 2 months; 3 African-American, 5 Hispanic, 1 white; mean standard score on TOWRE sight words: 85 (within mean) and TOWRE Phonemic Decoding: 83 (below mean)</p> <p>Control group: n = 9 (female = 5, male = 4); mean age 16 years, 1.5 months; 1 African-American, 7 Hispanic, 1 white; mean standard score on TOWRE sight words: 85 (within mean) and TOWRE Phonemic Decoding: 81 (below mean)</p>
Interventions	<p>Intervention: practice in phonemic awareness and decoding multi-syllable words using backwards chaining, followed by practice on Word Workout computer program (practice skills learned in teacher-instructed sessions)</p> <p>Control: not explicitly stated; however, probably treatment and schooling as usual throughout the training period (since all participants came from a remedial reading programme and participants were divided into intervention and control groups via random drawing (see p 48))</p> <p>Procedure: training was conducted by the researcher in small groups or one-to-one. Children did 3 × 15-minute sessions/week for 7 weeks</p>
Outcomes	<p>Time of post-test: immediately after training completed</p> <p>Primary outcomes: non-word reading accuracy (WJTA-III: Word Attack subtest; forms A and B), regular and irregular word reading accuracy (WJTA-III: Letter Word Identification subtest; forms A and B), non-word reading fluency (TOWRE: Phonemic Decoding Efficiency subtest (forms A and B)), regular and irregular word reading fluency</p>

	(TOWRE: Sight Word Efficiency subtest (forms A and B)) and reading comprehension (Gates-MacGinitie Reading Comprehension Subtest)	
Notes	<p>1. TOWRE sight words and TOWRE Phonemic Decoding standard scores calculated using raw scores given (see pp 771-2)</p> <p>2. 2 participants dropped out (1 from each group) and thus their pretest scores were removed. The thesis only provides information on the 18 participants who completed the training.</p> <p>3. Qualitative data (survey and focus groups) also collected</p> <p>Study start and end dates: not reported</p> <p>Funding: not reported</p> <p>Potential/declared conflicts of interest: none reported</p>	
Risk of bias		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Quote from publication: "students were randomly assigned to an experimental or control group by drawing names" (p 49)
Allocation concealment (selection bias)	Low risk	Comment: could not foresee assignment due to central allocation of participants to groups
Blinding of participants and personnel (performance bias) All outcomes	Unclear risk	Comment: no information provided; however, participants were adolescents with little understanding of reading treatment techniques and hence were unlikely to understand allocation
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Quote from publication: "Tests were administered by an experienced teacher who was not otherwise involved with the study ... to reduce tester bias" (p 49)
Incomplete outcome data (attrition bias) All outcomes	Low risk	Quote from publication: "One student from each group dropped out before the conclusion" (p 68) Comment: both groups experienced the same (low) dropout rate.
Selective reporting (reporting bias)	Unclear risk	Comment: data reported for all reading tests listed in methods; adequate detail for data to be included in analysis
Other bias	Low risk	Comment: none apparent

Hurford 1994

Methods	Randomised controlled trial (most likely) 1 intervention group (phonics + phoneme awareness) and 1 control group (untrained)
Participants	<p>Location/setting: mostly middle-class US elementary schools</p> <p>Criteria: standard score < 91 on the Word Attack test of the WRMT-R; standard score < 91 on the Word Identification subtests of the WRMT-R</p> <p>Recruits: children identified as at risk for RD with normal IQ or at risk for becoming poor readers with low IQ (GV poor readers)</p> <p>Sex: 48 male; 51 female</p> <p>Mean age: 80.35 months (SD and range not reported)</p> <p>Ethnicity: 92.8% white, 6% African-American, 5% Hispanic, and 7% Asian-American</p> <p>Sample size: 99 children</p> <p>Allocation: Half of the RDs and half of the GVs were included in the training group and the other half comprised the control groups. So, there were four groups (RD trained, GV trained, RD control and GV control). Group membership was determined by matching the students at risk for RD on the variables outlined in the method section and then RANDOMLY assigning them to either the T (treatment) or C (control) group. Statistical analysis was performed to determine that the T and C groups were equivalent” (quote from personal communication via email). Since this review did not use IQ as an exclusionary criteria, we merged the 2 trained groups (RD and GC) to form the intervention group, and merged the 2 untrained groups (RD and GV) to form the control group</p> <p>Intervention group: n = 49; mean = 25 (see notes below); 25 females and 24 males; mean age 79.8 months (SD and range not reported)</p> <p>Control group: n = 50; mean = 25 (see notes below); 26 females and 24 males; mean age 80.9 months (SD and range not reported)</p>
Interventions	<p>Intervention: intrasyllable discrimination training, phonemic blending and phonemic segmentation with letters. The training sequence was the same for each participant</p> <p>Control: no training</p> <p>Procedure: intervention was one-to-one, 15-20 minutes/session. Approximately 40 sessions - twice/week for approximately 20 weeks by computer and trainer</p>
Outcomes	<p>Time of post-test: < 1 month after training completed</p> <p>Primary outcomes: non-word reading accuracy (WJRMT-R: Word Attack subtest) and regular and irregular word reading accuracy (WJRMT-R: Word Identification subtest)</p>
Notes	<ol style="list-style-type: none"> 1. Study also included 332 children without reading difficulties, which we excluded as they did not meet the criteria for inclusion. 2. Dropouts for 486 participants initially screened: 55 (13.3%), “this loss in the participant pool due to attrition (13.3%) is similar to the attrition rate these school systems typically experience” (quote, p 649). 3. We used the Word Attack and Word Identification measures from the WRMT-R. Since we are including all poor readers regardless of IQ, we took the mean of the 2 untrained groups (RD and GV) for control data and the 2 trained groups (RD and GV) for experimental data. We also used the mean n for these groups, which was 25 in each case. 4. Contacted Hurford (20 September 2011) for means and SDs for primary outcomes (discrimination, segmentation, Word Identification and Word Attack measures) at pre- and post-test (supplied).

Hurford 1994 (Continued)

	Study start and end dates: not reported Funding: not reported Potential/declared conflicts of interest: none reported	
<i>Risk of bias</i>		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Quote from personal communication: "Group membership was determined by matching the students at risk for RD on the variables outlined in the method section and then RANDOMLY assigning them to either the T [treatment] or C [control] group."
Allocation concealment (selection bias)	Low risk	Comment: could not foresee assignment due to central allocation of participants to groups
Blinding of participants and personnel (performance bias) All outcomes	Unclear risk	Comment: no information provided; however, participants were children with little understanding of reading treatment techniques and hence were unlikely to understand allocation
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Personal communication: testing was done by someone who did not know the students.
Incomplete outcome data (attrition bias) All outcomes	Low risk	Quote from publication: "three groups lost approximately same percentage [13.3%] of participants" (p 649) Comment: all groups experienced the same (relatively low) dropout rate
Selective reporting (reporting bias)	Unclear risk	Comment: data reported for all reading tests listed in methods; adequate detail for data to be included in analysis
Other bias	Low risk	Comment: none apparent

Methods	<p>Randomised controlled trial</p> <p>2 intervention groups (phonics + phoneme awareness, Reading Recovery (not relevant)) and 2 control groups (untrained, Reading Recovery (not relevant))</p>
Participants	<p>Location/setting: year 2 children from English schools which provided Reading Recovery</p> <p>Criteria: 1 of 7 poorest year 2 scorers in 18 schools on the Diagnostic Survey (Clay 1985)</p> <p>Recruits: 42% received free school meals. 1 child was excluded from the study because of missing baseline data. All children had IQ in the mean range (92-96)</p> <p>Sex: 61% male; 39% female</p> <p>Mean age: not reported (SD not reported; range 6-6.6 years)</p> <p>Ethnicity: 16% spoke English as a second language</p> <p>Sample size: 142 children</p> <p>Allocation: random allocation (within schools) of poor readers to intervention and control groups</p> <p>Intervention groups:</p> <ol style="list-style-type: none"> 1. phonics + phoneme awareness: n = 96 (n = 92 for post-test data) (sex, mean age, SD, and range not reported) 2. Reading Recovery: n = 95 (n = 89 for post-test data) (sex, mean age, SD, and range not reported) <p>Control groups:</p> <ol style="list-style-type: none"> 1. untrained: n = 46 (n = 43 for post-test data) (sex, mean age, SD, and range not reported) 2. Reading Recovery: n = 41 (n = 40 for post-test data) (sex, mean age, SD, and range not reported)
Interventions	<p>Interventions:</p> <ol style="list-style-type: none"> 1. phonics + phoneme awareness (phonological training): “Following Bradley and Bryant (1985), this involved sound awareness training plus word building with plastic letters. The training initially focused on alliteration and rhyme but also included work on boundary sounds and vowels and digraphs in response to the child’s progress. Children also matched sounds with plastic letters and constructed words” (quote, p 234; phonics + phonological awareness). 2. Reading Recovery <p>Controls:</p> <ol style="list-style-type: none"> 1. untrained: children in within-school control groups received standard provision available in school. Since these children were poor readers, they received around 21 minutes of extra help per week with reading. 2. Reading Recovery <p>Procedure: intervention was 40 sessions (10 minutes each, one-to-one with tutor, spread over 7 months). 5 tutors delivered phonological training. Did not share details of intervention with classroom teachers</p>
Outcomes	<p>Time of post-test: immediately after training completed</p> <p>Primary outcomes: regular and irregular word reading accuracy (BAS word reading) and reading comprehension (Neale Prose Reading)</p>

Notes	<p>1. For the phonological training group, the article reported that the 6 poorest readers from 23 schools were allocated to either the phonological training (n = 4) or the within-school control (n = 2). This would equate to 92 participants in the phonological training group. However, Table 1 (p 232) reported that there were 96 participants in the phonological training group. We contacted Jane Hurry to explain this. We received a reply on the 16 January 2012: “I have now looked at the file and find that of the 23 Phon schools we actually selected the bottom 7 children from 5 of the schools. Of those 5 extra children, there was missing baseline data for 1, so that child never made it into the study. The other extra 4 were assigned to the intervention, hence the 96” (quote).</p> <p>2. We excluded the 22 Reading Recovery schools (and controls) from our analysis since it involved text reading (an exclusion criterion of our review).</p> <p>3. We excluded the 18 untrained control schools since the within-school controls were superior controls for the trained children because they were better matched for SES and learning environment.</p> <p>4. Contacted Hurry on 14 September about which subtests were used from the Neale Prose Reading. Replied that they used the accuracy and comprehension subtests to make up their Neale Prose Reading measure (see Table 1). We used this as a measure of reading comprehension.</p> <p>5. There were 3 post-tests: post-test 1 (after completion), post-test 2 (1 year later), post-test 3 (3.5 years later). We included the first post-test results in this review since all other studies in this review reported immediate post-test data.</p> <p>6. Contacted Hurry for clarification on:</p> <ul style="list-style-type: none"> i) participant numbers (Hurry responded on 16 January 2012; see above); ii) attrition (Hurry responded on 17 January 2012; see response in 'Risk of bias' table below), iii) which subtest of the Neale (Prose) Reading was used: Neale accuracy and comprehension scores (03 February 2012), iv) approximately how many minutes/hours the participants spent on phonological training per week (Hurry responded on 16 February 2012: “I confirm that each child was given 40 x 10 min individual sessions = 400 minutes” (quote)). <p>Study start and end dates: September 1992 to December 1996 Funding: “This work was conducted with the...funding of QCA” (quote, p 246) Declared/potential conflicts of interest: none reported</p>
-------	---

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Quote from publication: for the relevant groups (phonological training and within-school controls): the “six poorest readers randomly assigned to phonological training (N = 4) or to within-school control condition (N = 2)” (p 231)
Allocation concealment (selection bias)	Low risk	Comment: could not foresee assignment due to central allocation of participants to groups

Hurry 2007 (Continued)

Blinding of participants and personnel (performance bias) All outcomes	Unclear risk	Comment: no information provided; however, participants were children with little understanding of reading treatment techniques and hence were unlikely to understand allocation
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Quote from publication: “At each of the three post-tests, members of the research team tested the children ‘blind’, that is without knowing to which group children belonged” (p 233)
Incomplete outcome data (attrition bias) All outcomes	Low risk	Comment: for the relevant groups (phonological training and within-school controls), 4 and 3 (respectively) children dropped out between pre- and post-test 1. We requested more information from author. Received response on 17 January 2012 that some children “had failed to receive a sufficient amount of the intervention, usually as a result of moving school” while others “could not be tested because they had moved too far or were not traced” (quote). Thus, both groups experienced the same (relatively low) dropout rate for reasons extraneous to the study
Selective reporting (reporting bias)	Unclear risk	Comment: data reported for all reading tests listed in methods; adequate detail for data to be included in analysis
Other bias	Low risk	Comment: none apparent

Levy 1997

Methods	Randomised controlled trial 4 intervention groups (rime, onset, phoneme, whole word (not relevant)) and 1 control group (untrained)
Participants	Location/setting: grade 2 children from 16 schools in Canada Criteria: < 7 words read correctly on the WRMT Word Identification test; or < 7 words read correctly on the WRAT-R Word Identification test; or < 7 training words read correctly Recruits: 125 English-speaking children. Mean performance on WRMT at Grade 1.2 level and Word Identification subtest of WRAT-R scores in preschool range. On average only read 3 or 4 words from the set of 32 words to be trained Sex: not reported Mean age: not reported (SD not reported; range 5.9-7.9 years)

	<p>Ethnicity: not reported</p> <p>Sample size: 100 English-speaking children</p> <p>Allocation: children were randomly allocated to 5 groups: 4 intervention groups and 1 control group. 3 intervention groups did phonics training, so all these children were grouped together for the intervention group. The 4th group did whole word training (not relevant). The 5th (untrained) group was used as the control group</p> <p>Intervention groups:</p> <ol style="list-style-type: none"> 1. rime: n = 25 (sex, mean age, SD, and range not reported) 2. onset: n = 25 (sex, mean age, SD, and range not reported) 3. phoneme: n = 25 (sex, mean age, SD, and range not reported) 4. whole word: n = 25 (sex, mean age, SD, and range not reported) <p>Control group: n = 25 (sex, mean age, SD, and range not reported)</p>
<p>Interventions</p>	<p>Interventions:</p> <p>“The four training groups all learned to read the same set of 32 words, as well as participated in the classroom program... On each day of training, children in all groups read once only the entire set of 32 words printed on individual index cards. The groups differed in how the words were grouped during learning, and in the method of instruction” (quote, p 366)</p> <ol style="list-style-type: none"> 1. rime: “four written words of a rime family were shown together. First 15 days or until all 32 words pronounced correctly on 2 successive days: common rime segment for each family block was written in red to highlight the shared orthographic segment” (quote, p 366). Following 15 days or when criterion was met: “10 black and white trials where the child pronounced the 32 words printed in black ink once a day” (quote, p 368) 2. onset: “four written words per family block shared the initial consonant(s)-vowel segment” (quote, p 368). 15 colour trial days (or 2 successive correct readings): initial consonant(s)-vowel segment written in red. Following the 15 days or when criterion was met: maximum of 10 black and white trials (quote, p 368) 3. phoneme: “four written words for each block were randomly selected from the 32 words, with the restriction that no two onset or rime family members could be in the same block. The same eight random blocks were used on each day of training. There was no consistent relation among phonemic units in the four words, but for each word the letters of each phoneme were printed in a different colour... maximum of 15 colour trials and 10 black and white trials” (quote, p 368) 4. whole word: “four words per block randomly selected... words written in black ink... experiment read each word with no segmentation” (quote, p 368) <p>Control: received regular classroom regimen during the training phase</p> <p>Procedure: pre-test phase, training phase, post-training phase. One-to-one training</p>
<p>Outcomes</p>	<p>Time of post-test: immediately after training completed</p> <p>Primary outcomes: non-word reading accuracy (experimental: 48 new non-words) and regular word reading accuracy (experimental: 48 new regular words)</p>
<p>Notes</p>	<ol style="list-style-type: none"> 1. Paper presented 2 experiments. Experiment 1 focused on non-readers while experiment 2 focused on poor readers. Therefore, we only included experiment 2 in our review. 2. Intervention 4 (of experiment 2) trained irregular words and therefore we did not include this in our review or analysis.

3. Contacted author (B Levy) on 26 September 2011 for:

- i) mean age (and SDs) of participants: did not know;
- ii) number of males/females: did not know;
- iii) inclusion criteria: did not know;
- iv) details on the control group: same as the control group in experiment;
- v) length of training: depended on child's progress and speed of responding;
- vi) training group size: one-to-one

4. Since the rime, onset and phoneme training groups all trained phonics, we merged their results for the experimental data.

5. There were 2 measures that tested reading accuracy: non-words (onset non-words and rime non-words). We merged these 2 tests for a measure of reading accuracy: non-words. Similarly, there were 2 measures testing reading accuracy: regular words (onset words and rime words). We merged these 2 tests for 1 measure of reading accuracy: regular words.

6. There were 2 immediate post-tests: the day after completion, and 1 week after completion. We used the first post-test in this review.

Study start and end dates: not reported

Funding: "This research was supported by a grant to Betty Anne Levy from the Ontario Mental Health Foundation" (quote, p 386)

Declared/potential conflicts of interest: none reported

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Quote from publication: "twenty-five children were randomly assigned to each of the five training conditions" (p 378) Quote from personal communication: "children were randomly assigned to conditions as they arrived for the study, with the intention to keep numbers per condition as equal as possible in each school at all times. The idea was to balance for time of year effects and conditions in schools. Otherwise, assignment per condition was random and controlled by the tester."
Allocation concealment (selection bias)	Low risk	Comment: could not foresee assignment due to central allocation of participants to groups
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Quote from personal communication: "the teachers and parents knew the general purpose of the study but no details of manipulations, child assignments or individual child outcomes." Comment: participants were children with little understanding of reading treatment

Levy 1997 (Continued)

		techniques and hence were unlikely to understand allocation
Blinding of outcome assessment (detection bias) All outcomes	Unclear risk	Quote from personal communication: “the same testers scored all tests for both pre- and post-tests. No blinding of testers was attempted since the experimenters were largely the testers.” Comment: study used objective tests of literacy-related skills that are designed to avoid assessor bias
Incomplete outcome data (attrition bias) All outcomes	Low risk	Comment: no explicit information about attrition, but DF suggested all randomised participants were included in the analysis
Selective reporting (reporting bias)	Unclear risk	Comment: data reported for all outcome measures outlined in methods; adequate detail for data to be included in analysis
Other bias	Low risk	Comment: none apparent

Levy 1999

Methods	Randomised controlled trial (stratified randomisation) 2 intervention groups (onset rime + phoneme segmentation, whole word (not relevant)) and 1 control group (alternative training)
Participants	Location/setting: grade 2 classrooms of the Hamilton-Wentworth Roman Catholic Separate School Board, Canada Criteria: English speaker; score < 90 on the WRAT-3 Word Identification test; score > half a grade below appropriate grade on the WRMT Word Identification test; < 15 training words read correctly Recruits: 128 English-speaking children in Grade 2 Sex: 72 male; 56 female Mean age: 7 years 7 months (SD and range not reported) Ethnicity: mixed racial distributions Sample size: 96 English-speaking children Allocation: fast RAN and slow RAN poor readers randomly allocated to four groups (onset rime, phoneme segmentation, whole word, arithmetic). Two groups received phonics training (onset rime, phoneme segmentation), and so were merged. The arithmetic group was used as the control group Intervention groups: 1. onset rime + phoneme segmentation: n = 64 (sex, mean age, SD, and range not reported) 2. whole word: n = 32 (sex, mean age, SD, and range not reported) Control group: n = 32 (sex, mean age, SD, and range not reported)

Interventions	<p>Interventions:</p> <p>“On each day of training, children in all groups read through the set of 48 words once only. Each word was printed on a separate index card. On the 1st day only, the experimenter first read through the set only once, in a manner appropriate to modelling that training condition, and then the child read through the set in the same manner. On all subsequent days, the child read the words and the experimenter provided only corrective feedback. The critical differences among the three training conditions for the fast and the slow RAN groups were how the 48 words were grouped together during the presentation and how the words were segmented” (quote, pp 123-4)</p> <ol style="list-style-type: none"> 1. onset rime + phoneme segmentation: <ol style="list-style-type: none"> i) rime: “48 words were presented 4 at a time, where the word on each of the four cards presented together was from the same rime family and each was segmented by colouring the rime unit in red and the onset unit in black” (quote, p 124) ii) colour trials: 15 days or until criterion of entire 48 words read correctly on 2 successive days was met. Following the colour trials, the words were printed in black ink only. iii) phoneme: “Each phonemic unit was printed in a different colour for the 1st 15 days of training or until the criterion of two successive perfect readings was met” (quote). Following the colour trials, the words were printed in black and white. 2. whole word: “Each card contained a written word written in one of three colours.. each word was in a single colour and the experimenter pronounced the whole word with no segmental breaks. The child then read the whole words on each trial, with corrective feedback at the whole word level. On black and white trials, the colours were removed... all words were printed in black ink” (quote, p 124) <p>Control (arithmetic): “Help with addition and subtraction in one-on-one sessions” (quote, p 125)</p> <p>Procedure: all one-to-one training, outside of the classroom, for 15 minutes/day for 4 weeks</p>
Outcomes	<p>Time of test: day after completion of training: immediate</p> <p>Primary outcomes: non-word reading accuracy (experimental: 48 new non-words) and regular word reading accuracy (experimental: 48 new regular words)</p>
Notes	<ol style="list-style-type: none"> 1. While there were 6 intervention groups (fast and slow RAN rime, phoneme and whole word) our review focused on the rime and phoneme conditions since they were phonics training. 2. Since both the rime and phoneme intervention groups trained phonics, the experimental data used in this review was a mean of the fast and slow RAN rime and phoneme training groups (i.e. 4 groups). The control data was a mean of the fast and slow RAN control groups. 3. There were 2 immediate post-tests: the day after completion and 1 week after completion. We only used the first post-test in this review. <p>Study start and end dates: not reported</p> <p>Funding: “This research was supported by a grant to the first author from the Social Sciences and Humanities Research Council of Canada” (quote, p 115)</p> <p>Declared/potential conflicts of interest: none reported</p>
<p><i>Risk of bias</i></p>	

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	<p>Quote from publication: “the fastest RAN children were assigned to the four fast RAN training groups and the slowest RAN children were assigned to the four slow RAN training groups” (p 121)</p> <p>Comment: 1 fast RAN group and 1 slow RAN group were allocated to each type of training and a control group</p> <p>Quote from personal communication: “Children were randomly assigned to conditions as they arrived for the study, with the intention to keep numbers per condition as equal as possible in each school at all times. The idea was to balance for time of year effects and conditions in schools. Otherwise assignment per condition was random and controlled by the tester.”</p>
Allocation concealment (selection bias)	Low risk	<p>Comment: could not foresee assignment due to central allocation of participants to groups</p>
Blinding of participants and personnel (performance bias) All outcomes	Low risk	<p>Quote from personal communication: “the teachers and parents knew the general purpose of the study but no details of manipulations or child assignments or individual child outcomes.”</p> <p>Comment: participants were children with little understanding of reading treatment techniques and hence were unlikely to understand allocation</p>
Blinding of outcome assessment (detection bias) All outcomes	Unclear risk	<p>Quote from personal communication: “the same testers scored all tests for both pre and post tests. No blinding of testers was attempted since the experimenters were largely the testers.”</p> <p>Comment: study used objective tests of literacy-related skills that are designed to avoid assessor bias</p>
Incomplete outcome data (attrition bias) All outcomes	Low risk	<p>Comment: no explicit information about attrition, but analysis of number of children who met criterion after training suggests that all randomised participants were included in the analysis (i.e. 16 in each group)</p>

Levy 1999 (Continued)

Selective reporting (reporting bias)	Unclear risk	Comment: data reported for all outcome measures outlined in methods; adequate detail for data to be included in analysis
Other bias	Low risk	Comment: none apparent

Lovett 1990

Methods	Randomised controlled trial 2 intervention groups (phonics + sight words, sight words (not relevant)) and 1 control group (alternative training)
Participants	<p>Location/setting: children referred to the Learning Disabilities Research Program at The Hospital for Sick Children in Toronto, Canada</p> <p>Criteria: score < 25th percentile on 4 out of 5 reading tests (WRAT-3: Reading; WRMT-R: Word Identification; WRMT-R: Word Attack; Peabody Individual Achievement Test - Revised: Reading Recognition; GFW Sound-symbol Tests: Reading of Symbols); WISC-R Verbal and Performance IQ \geq 85; no English as second language, extreme hyperactivity, hearing impairment, brain damage, a chronic medical condition or serious emotional disturbance, attention deficits; aged 7-13 years</p> <p>Recruits: 54 disabled readers. WISC-R Mean Verbal IQ 98.4, SD 10.6; Mean Performance IQ 106.2, SD 12.6. Majority of participants were from families in the middle socioeconomic ranges according to the Blishen scales (Index $M = 43.6$, SD = 11.5, range 28.9-71.7)</p> <p>Sex: 38 male; 16 female</p> <p>Mean age: 8.4 years (SD 1.6; range 7-13 years)</p> <p>Ethnicity: not reported</p> <p>Sample size: 36 disabled readers</p> <p>Allocation: randomly assigned to 3 groups: REGEXC, REG=EXC, and control (CSS). This review used the REGEXC group as the intervention group and the CSS group was the control group (see notes for remaining group)</p> <p>Intervention groups:</p> <ol style="list-style-type: none"> 1. phonics + sight words: $n = 18^*$ (sex, mean age, SD, and range not reported) 2. sight words: $n = 18^*$ (sex, mean age, SD, and range not reported) <p>Control group: $n = 18^*$ (sex, mean age, SD, and range not reported)</p>
Interventions	<p>Interventions:</p> <ol style="list-style-type: none"> 1. phonics + sight words: REGEXC; "Regular words were taught by training the constituent letter-sound mappings. Exception words were introduced and rehearsed by whole-word methods alone... spelling training for regular words emphasized segmentation of the word into its individual <i>sounds</i>, with attention paid to the sequence of sounds, the sequence of individual letters, and any letter-sound patterns illustrated by the word" (quote, p 770-1) 2. sight words: "regular and irregular words taught the "exception word" way" (quote, p 770) <p>Control: CSS programme: problem solving and study skills training</p> <p>Procedure: 35 \times 1-hour sessions for each programme (4/week). Children instructed in pairs in special laboratory classrooms at a paediatric teaching hospital by special education</p>

	<p>teachers. “There was no attempt to explicitly control for other educational experiences of the children enrolled in these programs. Some were in special education placements in their community schools; some were not and had never been. For those subjects receiving any other individualized remedial instruction, their teacher was asked to refrain from training, rehearsing, or elaborating in any way on the instructional content the child was receiving as part of his or her experimental treatment program” (quote, p 771)</p>	
Outcomes	<p>Time of post-test: not stated explicitly but appears to be immediate Primary outcomes: regular word reading accuracy (experimental: trained and untrained words), irregular word reading accuracy (experimental: trained and untrained words), regular word reading fluency (experimental: trained and untrained words), irregular word reading fluency (experimental: trained and untrained words), regular word spelling (experimental: trained and untrained words), and irregular word spelling (experimental: trained and untrained words) Secondary outcomes: letter-sound knowledge (experimental: trained and untrained letter-sound rules)</p>	
Notes	<p>*Contacted Jan Frijters who supplied this information</p>	
Risk of bias		
Bias	Authors’ judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	<p>Quote from publication: “children were randomly assigned to a treatment condition and to a particular teacher” (p 771) Quote from personal communication: “children were matched on decoding ability and then random number tables were used to randomly assign treatment to pair and to assign teacher to pair.” Comment: best described as matching with randomisation.</p>
Allocation concealment (selection bias)	Low risk	<p>Quote from personal communication: “the PI assigned treatments and teachers to child pair based on participant identity alone. Neither children nor teachers would have had contact with the person doing the assignment, as all contact prior to this point was with study psychometrists.” Comment: could not foresee assignment due to central allocation of participants to groups</p>
Blinding of participants and personnel (performance bias) All outcomes	Unclear risk	<p>Quote from personal communication: “since this is a verbally-administered intervention with quite explicit and structured content, and teachers were trained on the</p>

		materials used, teachers could not be blind to the particular treatment they were teaching. Participants were not told what their assignments were, but on consent forms were told that they would participate in one of three conditions, with all conditions described. Teachers did not reveal condition to participants.”
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Quote from personal communication: “All standardized/norm referenced assessments were administered by trained psychometrists who were blind to assignment; however, some content-related and experimental measures (e.g. the four word lists) were administered by teachers themselves at the pre-specified testing intervals. In the former case, psychometrists would have had the participants name and testing folder alone, not the master subject-list.”
Incomplete outcome data (attrition bias) All outcomes	Low risk	Comment: in the publication, there was no explicit information about attrition; the fact that DF varied between tests suggests missing data for some children for some tests Quote from personal communication: “this one has puzzled us. We would typically report dropouts and/or discontinuations. Given the design, we would have expected a df of 50, which is what is reported for most measures. The lower df would likely indicate not dropped-out participants, but equipment errors, basal/ceiling problems, etc. that may have invalidated particular tests, or in the case of speed specifically (reported as 41 df) a failure of the voice onset recording device.” Comment: given that equipment errors etc. occur on a random basis, the lower DF were unlikely to relate to bias
Selective reporting (reporting bias)	Unclear risk	Comment: data reported for all outcome measures outlined in methods; adequate detail for data to be included in analysis
Other bias	Low risk	Comment: none apparent

Methods	Randomised controlled trial 2 intervention programmes (phonics + phoneme awareness, word identification (not relevant)) and a control (alternative training)
Participants	<p>Location/setting: children referred to the Clinical Unit at The Hospital for Sick Children in Toronto, Canada</p> <p>Criteria: score < 25th percentile on 4 out of 5 reading tests (WRAT-3: Reading; WRMT-R: Word Identification; WRMT-R: Word Attack; Peabody Individual Achievement Test-Revised: Reading Recognition; GFW Sound-symbol Tests: Reading of Symbols); WISC-R Verbal and Performance IQ ≥ 85; no English as second language, extreme hyperactivity, hearing impairment, brain damage, a chronic medical condition or serious emotional disturbance, attention deficits; aged 7-13 years</p> <p>Recruits: 166 reading disabled children. Mean IQ on WISC-3 or WISC-R: Verbal IQ M 92, SD 13.7, Performance IQ M 98.7, SD 14.3. On average, sample > 2 SD below age-norm expectations at referral, with half of the children consistently below the first percentile for age on standardised achievement measures. Of these 166, 84.3% of the sample (140 participants) could be classified into 1 of 3 subgroups: 54.3% double deficit, 22.1% phonological deficit, 23.6% visual naming-speed deficit</p> <p>Sex: 113 males; 53 females</p> <p>Mean age: 9.9 (SD 1.6 years; range 7-13 years)</p> <p>Ethnicity: not reported</p> <p>Sample size: 88 reading disabled children</p> <p>Allocation: 140 children randomly assigned to 1 of 3 treatments: PhAB training; WIST Program (not relevant to this review); and CSS (controls). In this review, the PhAB trainees were the intervention group and the CSS were the control group</p> <p>Intervention groups:</p> <ol style="list-style-type: none"> 1. phonics + phoneme awareness: n = 51 (sex, mean age, SD, and range not reported) 2. word identification: n = 52 (sex, mean age, SD, and range not reported) <p>Control group: n = 37 (sex, mean age, SD, range not reported)</p>
Interventions	<p>Interventions:</p> <ol style="list-style-type: none"> 1. phonics + phoneme awareness: PhAB skills were trained with oral and written presentations of letter-sound and letter-cluster-sound correspondences. Word segmenting and blending, sound segmentation and blending, rhyming. Special orthography used to teach letter sounds: “the special orthography is a temporary convention used to highlight salient features of some letters; it provides visual cues to the child with RD such as symbols over long vowels (macrons), letter size variation, and connected letters to facilitate initial learning” (quote, p 337) 2. word identification: “instruct children in the acquisition, use, and monitoring of different word identification strategies” (quote, p 338) <p>Control: the CSS Program taught organisational strategies, academic problem solving, study and self-help techniques. Children in the CSS programme received the same amount of individualised teacher attention as did children in the remedial reading programmes</p> <p>Procedure: children received 35 hours of instruction (1-hour sessions, 4 times/week) on a 2:1 or 3:1 ratio in special laboratory classrooms at a paediatric teaching hospital or in affiliated schools in the Toronto metropolitan area</p>

Outcomes	<p>Time of post-test: immediately after training completed</p> <p>Primary outcomes: non-word reading accuracy (WJRMT: Word Attack subtest), regular word reading accuracy (experimental: 149 untrained regular words), and irregular word reading accuracy (experimental: 149 untrained exception words)</p> <p>Secondary outcomes: phoneme awareness (GFW Sound Symbol Tests: Sound Analysis subtest)</p>	
Notes	<ol style="list-style-type: none"> 1. Contacted Frijters (on 4 October 2011) about means and SDs for reading measures from each of the 3 training conditions. We received an Excel file with means and SDs. 2. Asked whether there was an overlap in participants across 1994, 1997, and 2000 papers published by their laboratory (n = 62 in 1994 paper, n = 122 in 1997 paper, and n = 166 in 2000 paper). It was confirmed that there was an overlap in participants between the papers. Therefore, we decided to only include the 2000 paper for this review to limit any over representation of the data in the final meta-analysis. 3. The second intervention group did the WIST Program. The WIST contained > 2 training components (word identification by analogy, seeking the part of the word that you know, attempting variable vowel pronunciations, 'peeling off' prefixes and suffixes in a multi-syllabic word) and so was not included in this review. 4. 2 measures tested Reading Accuracy: non-words (GFW: Reading of Symbols and WJRMT-R: Word Attack). We included the WJRMT-R as it is a very widely used test with known reliability. 5. There were multiple measures of phoneme awareness. We selected GFW sound analysis because it was well matched between groups before training. <p>Study start and end dates: not reported</p> <p>Funding: "This article was supported by operating grants to Dr Lovett from the Ontario Mental Health Foundation, the Velleman Foundation, and the Social Sciences and Humanities Research Council of Canada. Additional support for data analysis and manuscript preparation was provided by a Shannon Award to Dr Lovett and to Drs. Robin Morris and Maryanne Wolf from the National Institute of Child Health and Human Development and further supported by NICHD award No. 1 RO1 HD30970-01 A2 to the same investigators" (quote, p 355)</p> <p>Declared/potential conflicts of interest: none reported</p>	
<i>Risk of bias</i>		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	<p>Quote from publication: "the experimental design in which the original 166 children participated involved random assignment to one of three active treatment programs" (p 336)</p> <p>Quote from personal communication: "children were matched on decoding ability and then random number tables were used to randomly assign treatment to pair and to assign teacher to pair."</p>

		Comment: best described as matching with randomisation.
Allocation concealment (selection bias)	Low risk	Quote from personal communication: “the PI assigned treatments and teachers to child pair based on participant identity alone. Neither children nor teachers would have had contact with the person doing the assignment, as all contact prior to this point was with study psychometrists.” Comment: could not foresee assignment due to central allocation of participants to groups
Blinding of participants and personnel (performance bias) All outcomes	Unclear risk	Quote from personal communication: “since this is a verbally-administered intervention with quite explicit and structured content, and teachers were trained on the materials used, teachers could not be blind to the particular treatment they were teaching. Participants were not told what their assignments were, but on consent forms were told that they would participate in one of three conditions, with all conditions described. Teachers did not reveal condition to participants.” Comment: participants were children with little understanding of reading treatment techniques and hence were unlikely to understand allocation
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Quote from personal communication: “all standardized/norm referenced assessments were administered by trained psychometrists who were blind to assignment; however, some content-related and experimental measures (e.g. the four word lists) were administered by teachers themselves at the pre-specified testing intervals. In the former case, psychometrists would have had the participants name and testing folder alone, not the master subject-list.”
Incomplete outcome data (attrition bias) All outcomes	Low risk	Comment: extra data provided by author revealed that the data of all randomised participants were included in the analyses
Selective reporting (reporting bias)	Unclear risk	Comment: data reported for all outcome measures outlined in methods; adequate

		detail for data to be included in analysis
Other bias	Low risk	Comment: none apparent

McArthur 2015a

Methods	<p>Quasi-randomised controlled trial</p> <p>3 treatment groups (phonics, sight words (not relevant), mixed (not relevant)) and 3 control groups (no-training double-baseline period for phonics group, sight words (not relevant), mixed (not relevant))</p>
Participants	<p>Location/setting: Sydney, Australia</p> <p>Criteria: scored below the mean range for their age (i.e. had a Z score lower than -1) on the CC2 irregular-word reading test or non-word reading test. No history of neurological or sensory impairment as indicated on a background questionnaire. Used English as their primary language at school and at home</p> <p>Recruits: full study included 141 dyslexic children recruited from schools, clinics, and newspaper advertisements. This review included the 39 participants who completed 8 weeks of no training (control) and then 8 weeks of pure phonics training (intervention)</p> <p>Sex: 63.8% male; 36.2% female</p> <p>Mean age: 9.42 years (SD 1.71; range 7-12 years)</p> <p>Ethnicity: not reported</p> <p>Sample size: 39 dyslexic children</p> <p>Allocation: quasi-randomised allocation procedure. Full study had 3 recruitment periods. 3 groups were recruited in each recruitment period. The children included in this review were recruited in the first recruitment period (months 1-6). The other children, recruited for the 2nd and 3rd groups in 2nd and 3rd periods, were not included in this review since they did a mixture of phonics + sight word training). A between-groups ANOVA established that the groups did not differ in reading ability or age prior to training</p> <p>Intervention groups:</p> <ol style="list-style-type: none"> 1. phonics: n = 39 (sex, mean age, SD, and range not reported) 2. sight words: n = 40 (sex, mean age, SD, and range not reported) 3. mixed: n = 38 (sex, mean age, SD, and range not reported) <p>Control groups:</p> <ol style="list-style-type: none"> 1. phonics T1: n = 39 (sex, mean age, SD, and range not reported) 2. sight words: n = 40 (sex, mean age, SD, and range not reported) 3. mixed: n = 38 (sex, mean age, SD, and range not reported)
Interventions	<p>Interventions:</p> <ol style="list-style-type: none"> 1. phonics: children were instructed to do the phonics training at home for 30 minutes/day, 5 days/week, for 8 weeks. All training was done on a computer using a modified version of the Lexia® Strategies for Older Students, which uses a wide variety of games and exercises to teach the pairing of written stimuli (i.e. letters, letter clusters, syllables, morphemes, whole words, phrases, and sentences) to the spoken versions of those stimuli. The modified programme thus focused on training GPCs either alone, within parts of words (i.e. syllables), or within regular words. Phonics training focused on accuracy rather than fluency.

	<p>2. sight words: children were taught to read irregular words by sight using the DingoBingo game.</p> <p>3. mixed: children did both phonics and sight word training, alternating from day to day.</p> <p>Controls:</p> <ol style="list-style-type: none"> 1. phonics T1: prior to training, children completed a double-baseline period with outcome measures tested before and after 8 weeks of no training. 2. sight words: prior to training, children completed a double-baseline period with outcome measures tested before and after 8 weeks of no training. 3. mixed: prior to training, children completed a double-baseline period with outcome measures tested before and after 8 weeks of no training.
<p>Outcomes</p>	<p>Time of post-test: immediately after no-training period (control) and then immediately after 8 weeks of phonics training (experimental)</p> <p>Primary outcomes: trained and untrained irregular word reading accuracy and non-word reading accuracy</p> <p>Secondary outcomes: word and non-word reading fluency and reading comprehension</p> <p>Relevant measures: non-word accuracy (experimental: 20 untrained non-words printed on flashcards), irregular words (trained) accuracy (experimental: 30 flashcards), irregular words (untrained) accuracy (experimental: 30 flashcards), non-word fluency (TOWRE: Non-word subtest), mixed/regular word fluency (TOWRE: Sight Word subtest), reading comprehension (Test of Everyday Reading Comprehension)</p>
<p>Notes</p>	<ol style="list-style-type: none"> 1. In addition to the phonics groups, 2 groups in this study did phonics + sight word training. Since “this review was focused on phonics training, we included data on the ”purest“ example of this - i.e. gains in outcome measures in Group 1 before and after they did 8 weeks of phonics, and the we compared those gains to control data from the same group of children - i.e. gains in the same outcomes measures in Group 1 before and after an 8-week no training period” (quote from personal communication with author). 2. It is noteworthy that although all children were tested for their non-verbal intelligence, children with non-verbal IQ scores below the mean range were not excluded from the study since intelligence does not appear to predict reading ability or response to treatment. 3. Contacted author for the numbers for sex of participants. <p>Study start and end dates: not reported</p> <p>Funding: “This research was funded by NHMRC Project 488518 and ARC DP0879556” (quote, p 406)</p> <p>Declared/potential conflict of interest: “no potential conflicts of interest with respect to the research, authorship, and/or publication of this article” (quote, p 406)</p>
<p><i>Risk of bias</i></p>	
<p>Bias</p>	<p>Authors’ judgement</p>
<p>Random sequence generation (selection bias)</p>	<p>Low risk</p> <p>Support for judgement</p> <p>Comment: as noted in the study, this was a quasi-randomised controlled trial</p> <p>Quote from publication: “There is good evidence that this quasi-randomised alloca-</p>

		<p>tion procedure did not bias the outcomes of this study. First, the groups were very well matched prior to training (see Table 1). Second, for all but one outcome, groups made similar gains after 16 weeks of training, indicating that allocation did not produce any group that was unusually responsive or unresponsive to treatment. Third, for the exceptional outcome, the group difference was in the predicted direction, indicating that superior group performance was a result of a genuine experimental effect rather than a group allocation effect. Fourth, this study was designed so that there could be no possible bias between allocation to intervention and control groups since each individual participated in both control and intervention periods, and any gains in the control period were controlled for in the intervention period statistically (i.e. we used a double-baseline design that gauged the effect of no training in each and every participant before they did training)” (p 398)</p>
<p>Allocation concealment (selection bias)</p>	<p>Low risk</p>	<p>Quote from publication: “Each recruitment period had a fixed start date and an end date. Children were allocated to their group according to when they were recruited for the study. Since children could be allocated to only one group, it is highly unlikely that lack of allocation concealment introduced bias into the study” (pp 398-9)</p>
<p>Blinding of participants and personnel (performance bias) All outcomes</p>	<p>Low risk</p>	<p>Quotes from publication: “Unlike drug trials, cognitive treatment trials find it difficult to guarantee double blinding because the type of training cannot be completely concealed from a volunteer. However, neither parents nor children were told their group allocation, and it is highly unlikely that they had the expertise to ascertain the type of training that they were receiving (i.e. they were blind to group allocation). Furthermore, all children received exactly the same type of training in this study. The only difference was the order in which they did the training. This would further obscure group allocation to children and their parents” (p 399)</p>

McArthur 2015a (Continued)

		“...we employed four casual testers to help two principal testers. With careful planning, we ensured that no tester assessed the same child twice, and no tester was aware of the child’s group allocation (i.e. the tester was blind to group allocation)” (p 399)
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Quote from publication: “...we employed four casual testers to help two principal testers. With careful planning, we ensured that no tester assessed the same child twice, and no tester was aware of the child’s group allocation (i.e. the tester was blind to group allocation)” (p 399)
Incomplete outcome data (attrition bias) All outcomes	Low risk	Comment: 37 participants dropped out in total (26%). There were similar numbers in each group, and reasons for dropout were random. This is similar to McArthur 2015b , which used almost identical methods. This suggests that attrition was not unusual for reading training studies of this type, and is similar to mean attrition rates for cognitive behavioural interventions done with children with clinical problems (Karlson 2009).
Selective reporting (reporting bias)	Unclear risk	Comment: data reported for all outcome measures outlined in methods; adequate detail for data to be included in analysis
Other bias	Low risk	Comment: none apparent

McArthur 2015b

Methods	Randomised controlled trial (minimisation) 2 treatment groups (phonics, sight words (not relevant)) and 2 control groups (no-training double-baseline period for phonics group, double-baseline for sight words group (not relevant))
Participants	Location/setting: Sydney, Australia Criteria: scored below the mean range for their age (i.e. had a Z score lower than -1) on the CC2 irregular-word reading test or non-word reading test. No history of neurological or sensory impairment as indicated on a background questionnaire. Used English as their primary language at school and at home Recruits: full trial included 85 dyslexic children recruited from the community. The group included in this review - the phonics group - comprised 46 participants Sex: 46.3% male; 53.7% female

	<p>Mean age: group 1: 9.53 years (SD 1.51; range 7-12 years); group 2: 9.58 years (SD 1.45; range 7-12 years)</p> <p>Ethnicity: not reported</p> <p>Sample size: 46 dyslexic children</p> <p>Allocation: children were allocated to groups using minimisation randomisation (balanced 1:1 for age, CC2 non-word reading, CC2 irregular word reading; executed using MINIMPY; Saghaei 2011) (see p 10).</p> <p>Intervention groups:</p> <ol style="list-style-type: none"> 1. phonics: n = 46 (sex, mean age, SD, and range not reported) 2. sight words: n = 53 (sex, mean age, SD, and range not reported) <p>Control groups:</p> <ol style="list-style-type: none"> 1. phonics T1: n = 46 (sex, mean age, SD, and range not reported) 2. sight words: n = 53 (sex, mean age, SD, and range not reported)
Interventions	<p>Interventions:</p> <ol style="list-style-type: none"> 1. phonics: phonics training administered 5 days/week, 30 minutes/day, for 8 weeks, using an online reading training program called LiteracyPlanet. It taught “phonics using 9 exercises across 220 levels that increased in difficulty to train the explicit phonological decoding and encoding of consonants, short vowels, long vowels, blends, digraphs, the bossy e rule, plurals, soft ‘c’ and ‘g,’ diphthongs, ‘r’ sounds, and silent letters. No exercises included irregular words, sentences, or paragraphs of text” (quote p 8). 100% accuracy was required to move to the next level. 2. sight words: children were taught to read irregular words by sight using the exercises in LiteracyPlanet. <p>Controls:</p> <ol style="list-style-type: none"> 1. phonics T1: prior to training, children completed a double-baseline period with outcome measures tested before and after 8 weeks of no training. 2. sight words: prior to training, children completed a double-baseline period with outcome measures tested before and after 8 weeks of no training.
Outcomes	<p>Time of post-test: immediately after no-training period (control) and then immediately after 8 weeks of phonics training (experimental)</p> <p>Primary outcomes: trained and untrained irregular word reading accuracy and non-word reading accuracy</p> <p>Secondary outcomes: word and non-word reading fluency and reading comprehension</p> <p>Relevant measures: trained and untrained irregular words (experimental: 58 flash cards), non-word reading accuracy (experimental: 39 untrained non-words); non-word reading fluency (TOWRE: non-word subtest), mixed/regular word reading fluency (TOWRE: sight word subtest), reading comprehension (Test of Everyday Reading Comprehension)</p>
Notes	<ol style="list-style-type: none"> 1. McArthur 2015b was a replication of McArthur 2015a except the former was randomised while the latter was pseudorandomised, and the former included 2 groups and the latter included 3 groups. 2. Quote from personal communication with author: “In addition to the phonics groups, one group in this study did phonics + sight word training”. Since “this review was focused on phonics training, we included data on the ‘purest’ example of this - i.e. gains in outcome measures in group 1 before and after they did 8 weeks of phonics, and the we compared those gains to control data from the same group of children - i.e. gains in the same outcomes measures in group1 before and after an 8-week no training

period.”

3. It is noteworthy that although all children were tested for their non-verbal intelligence, children with non-verbal IQ scores below the mean range were not excluded from the study since intelligence does not appear to predict reading ability or response to treatment.

4. Contacted author for the numbers for sex of participants.

Study start and end dates: January 2011 to December 2013 (see p 4)

Funding: “This research was funded by NHMRC Project 488518 and ARC DP0879556” (quote, p 19)

Declarations/potential conflicts of interest: “The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript” (quote, p 19). At the time of publication, Associate Professor Genevieve McArthur was an Academic Editor of PeerJ, which may be considered a competing interest

<i>Risk of bias</i>		
Bias	Authors’ judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Quote from publication: “Children were allocated to groups using minimisation randomisation (balanced 1:1 for age, CC2 nonword reading, CC2 irregular word reading; executed using MINIMPY; Saghaei, 2011), which is considered the most appropriate sequence allocation procedure for trials comprising fewer than 100 participants. It is considered methodologically equivalent to randomisation by CONSORT” (p 10)
Allocation concealment (selection bias)	Low risk	Quote from publication: “The lead research assistant on the project allocated children to each group and arranged their training. They concealed group allocation from research assistants who conducted the test session” (p 10)
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Quote from publication: “Unlike drug trials, it is difficult to guarantee double blinding in cognitive treatment studies. However, parents and children were not told their group allocation, and all children received exactly the same type of training (in different orders). Most parents and children lack the expertise to discriminate between different types of reading. In addition, no tester assessed the same child twice, and no tester was aware of the child’s group allocation (i.e. the tester was blind to

McArthur 2015b (Continued)

		group allocation). Thus, it is highly likely this study used a double-blind procedure” (p 11)
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Quote from publication: “In addition, no tester assessed the same child twice, and no tester was aware of the child’s group allocation (i.e. the tester was blind to group allocation)” (p 11)
Incomplete outcome data (attrition bias) All outcomes	Low risk	Comment: 35 participants in total dropped out (29%). There were similar numbers in each group, and reasons for dropout were random. This is similar to McArthur 2015a , which used almost identical methods. This suggests that attrition was not unusual for reading training studies of this type, and is similar to mean attrition rates for cognitive behavioural interventions done with children with clinical problems (Karlson 2009) .
Selective reporting (reporting bias)	Unclear risk	Comment: data reported for all outcome measures outlined in methods; adequate detail for data to be included in analysis
Other bias	Low risk	Comment: none apparent

Savage 2003

Methods	Randomised controlled trial 3 intervention groups (phonics + phonemes, phonics + rimes, phonics + mixed) and 1 control group (untrained)
Participants	Location/setting: 9 schools in the London Borough of Sutton, UK Criteria: 108 year 1 children across 9 schools with the lowest scores on screening tests for phonological awareness (nursery rhymes, rhyme matching, rhyme generation, blending, segmentation) and reading (nonsense word reading, word reading and spelling, letter-sound knowledge); English speaking Recruits: 108 English-speaking readers in year 1 were selected. Sex: 64 males; 44 females Mean age: 5 years 9 months (SD not reported; range 5 years 0 months to 6 years 3 months) Ethnicity: not reported Sample size: 104 year 1 children Allocation: “within each school, children were allocated to an intervention condition (usually nine children) or to a control condition (usually three children)” (quote, p 219)

	<p>. Personal communication: "this was done using an (online) random number generator set with parameters 1-4, for each school allowing placing into each of the interventions. ..Child-level allocation to intervention versus control within each school was again undertaken using random number generator" (quote)</p> <p>Intervention groups:</p> <ol style="list-style-type: none"> 1. phonics + phonemes: n = 26 (sex, mean age, SD, and range not reported) 2. phonics + rimes: n = 26 (sex, mean age, SD, and range not reported) 3. phonics + mixed: n = 26 (sex, mean age, SD, and range not reported) <p>Control group: n = 26 (sex, mean age, SD, and range not reported)</p>
Interventions	<p>Interventions:</p> <p>"... in each session, all children started with letter-sound learning activities using a range of multi-sensory approaches (e.g. saying, looking, tracing) to learn letter sounds supported by the Jolly Phonics stories and actions" (quote, p 53); and "principles of segmenting and blending with a limited number of sounds" (quote, p 53). This was followed by 10-minutes of training on phonemes (for the phoneme training group), on rimes (for the rime training group) or on both (for the mixed training group). This, in turn, was followed by 5 minutes of phonological awareness training: "games tailored to phonemes or rhymes respectively" (quote, p 53). From this point in each session, the training varied between intervention groups</p> <ol style="list-style-type: none"> 1. phonics + phonemes: trained with <i>SoundWorks</i>: an 'a-board'; writing on lines (with 'slips' and 'foldovers': cards with vowel markers or spaces to write vowels); 'spelling from your head'; 'read the word'; and 'sound it out' with an adult. 2. phonics + rimes: practiced rimes with plastic letters along with writing words, simple word searches, using onset rime 'word fans', sorting words into '-an' and '-at' groups and using onset sound frames (depicted as elements in a picture of a caterpillar's body). 3. phonics + mixed: did a mixture of the 2 interventions above along with analysing words using their phonemic elements (e.g. 'at' made up of 'a' and 't') and using phonemes and rimes in word building. <p>Control: "children remained in class and undertook the word-level work appropriate to the second term of Year 1 of the National Literacy Strategy in their normal fashion" (quote, p 55)</p> <p>Procedure: LSAs conducted training in small groups (typically 4 children per group - as per email from Savage on 30 November 2011). 20-minute sessions, 4 times/week, for a period of 9 weeks at school</p>
Outcomes	<p>Time of post-test: not stated explicitly but appeared to be immediate.</p> <p>Primary outcomes: non-word reading accuracy (experimental: high rime non-words and low rime non-words), regular word reading accuracy (experimental: 6 regular words), regular word spelling (experimental: 6 regular words), letter-sound knowledge (experimental: "two sets of cards each containing 13 of the 26 letters of the alphabet presented one letter per card" (quote, p 218)), and phoneme awareness (experimental: onset-rime segmentation)</p>
Notes	<ol style="list-style-type: none"> 1. Similar design to Savage 2003 but done on a new sample of the same size (personal communication from Robert Savage on 30 November 2011) 2. Contacted Savage about: <ol style="list-style-type: none"> i) dropouts (on 24 January 2012): 4 dropouts, 1 from each group

ii) training group size (on 11 February 2012): typically 4 in each training group

3. Since the 3 intervention groups all consisted of phonics and phonological awareness training, we have used the combined mean scores (and SDs) at pre- and post-tests (see Table 3, p 222).

4. 2 tests used to measure reading accuracy: non-words (high rime non-words and low rime non-words). These 2 tests were normed.

5. 3 tests used to measure phoneme awareness: rime matching, onset-rime segmentation, and phoneme segmentation. We included the onset-rime segmentation as its intervention and control pretest scores had the best match.

Study start and end dates: not reported

Funding: financial support provided by the JJ Trust and the Helen Arkell Dyslexia Association

Declared/potential conflicts of interest: none reported

<i>Risk of bias</i>		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	<p>Quote from publication: "within each school, children were allocated to an intervention condition (usually nine children) or to a control condition (usually three children). Schools themselves decided on the precise composition of each of the sub-groups of three to four children who went together with an LSA for each intervention session based upon their knowledge of the children's social networks, so intervention groups varied slightly in size across schools" (p 219)</p> <p>Quote from personal communication: "this was done using an (online) random number generator set with parameters 1-4, for each school allowing placing into each of the interventions. Schools decided on suitability of children for intervention (as we note on page 219), though only 1 child was removed on teacher request. Child-level allocation to intervention versus control within each school was again undertaken using random number generator. However schools decided the precise composition of (the already selected) intervention child groups to create groups of children who got on well" (Savage 2003).</p> <p>Quote from personal communication: "the allocation was random at school and student-level. The composition of small</p>

		groups of children WITHIN the allocated random conditions was (and I recall, was very occasionally) adjusted only on the suggestion of classroom teachers to make the groups more functional at the social level (an e.g. I recall is a particular group of 4 randomly-allocated kids which included 3 'noisy' boys and a very shy girl), thus we might move the groups a bit for the delivery of the intervention. The initial randomisation was always respected. It was to avoid major problems that we would do this rather than to find groups who particularly got on, hence it was rare this happened. The key point is that the initial randomisation of condition was always intact, the grouping for the purpose of intervention delivery was occasionally adjusted" (Savage 2003).
Allocation concealment (selection bias)	Low risk	Quote from personal communication: "I did this allocation independent of those running the study and of co-author(s) Carless and Stuart. Carless led the TA training, so I judge allocation to be concealed, and not possible to predict" Comment: could not foresee assignment due to central allocation of participants to groups
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Quote from publication: "teachers were told who the control children and intervention children were, and were also reinforced at training and during the intervention to treat the control children in the same way as they would if no intervention was taking place for other children" (p 221) Comment: no information provided; however, participants were children with little understanding of reading treatment techniques and hence were unlikely to understand allocation
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Quote from personal communication: "Pre-testing was undertaken as a screen of all children in schools before we identified and allocated the 'at-risk readers', (see Consort flow diagrams in both papers) so in this sense it is entirely blind... There was no blinding of post-testing in relation

Savage 2003 (Continued)

		to the intervention condition as TAs did both (though see comments above on the 3 horse race). However classroom assistants also did not know of the theoretical contrasts (and they were definitely blind to the status of the high-rime and low-rime non-words in the 2003 study as these were randomised as a set of 12 items for pre-testing and post-testing). TAs were not told at any point of any research predictions regarding the relationship between intervention and outcome (e.g. hypothesis of possible link between phoneme-based intervention and raise phoneme awareness at post-test, and similar for rimes etc.).”
Incomplete outcome data (attrition bias) All outcomes	Low risk	Comment: 4 dropouts - 1 in each of the 4 groups
Selective reporting (reporting bias)	Unclear risk	Comment: data reported for all outcome measures outlined in methods; adequate detail for data to be included in analysis
Other bias	Low risk	Comment: none apparent

Savage 2005

Methods	Randomised controlled trial 3 intervention groups (phonics + phonemes, phonics + rhymes, phonics + mixed) and 1 control group (untrained)
Participants	<p>Location/setting: 9 schools in the London Borough of Sutton, UK</p> <p>Criteria: 108 year 1 children across 9 schools with the lowest scores on screening tests for phonological awareness (nursery rhymes, rhyme matching, rhyme generation, blending, segmentation) and reading (nonsense word reading, word reading and spelling, letter-sound knowledge); English speaking</p> <p>Recruits: 108 English-speaking readers in year 1 were selected.</p> <p>Sex: 54 males and 54 females</p> <p>Mean age: not reported</p> <p>Ethnicity: not reported</p> <p>Sample size: 52 year 1 children</p> <p>Allocation: the same as Savage 2003. That is random allocation of schools to 1 of 4 groups: 3 intervention groups (1 doing phoneme training, 1 doing rhyme training, and 1 doing a mix of both) and 1 control group (untrained). And then random allocation of children to treatment and control groups within schools. Since the 3 interventions trained phonics and phonological awareness, their data were merged for the Intervention group</p> <p>Intervention groups:</p>

	<ol style="list-style-type: none"> 1. phonics + phonemes: n = 26 (sex, mean age, SD, and range not reported) 2. phonics + rhymes: n = 26 (sex, mean age, SD, and range not reported) 3. phonics + mixed: n = 26 (sex, mean age, SD, and range not reported) <p>Control group: n = 26 (sex, mean age, SD, and range not reported)</p>
<p>Interventions</p>	<p>Interventions:</p> <p>”In each session, all children started with letter-sound learning activities using a range of multi-sensory approaches (e.g. saying, looking, tracing) to learn letter sounds supported by the Jolly Phonics stories and actions“ (quote, p 53); and ”principles of segmenting and blending with a limited number of sounds“ (quote, p 53). This was followed by 10-minutes of training on phonemes (for the phoneme training group), on rhymes (for the rhyme training group) or on both (for the mixed training group). This, in turn, was followed by 5 minutes of phonological awareness training: ”games tailored to phonemes or rhymes respectively“ (quote, p 53). From this point in each session, the training varied between intervention groups</p> <ol style="list-style-type: none"> 1. phonics + phonemes: trained with <i>SoundWorks</i>: an ‘a-board’; writing on lines (with ‘slips’ and ‘foldovers’: cards with vowel markers or spaces to write vowels); ‘spelling from your head’; ‘read the word’; and ‘sound it out’ with an adult. 2. phonics + rhymes: practiced rhymes with plastic letters along with writing words, simple word searches, using onset rhyme ‘word fans’, sorting words into ‘-an’ and ‘-at’ groups and using onset sound frames (depicted as elements in a picture of a caterpillar’s body). 3. phonics + mixed: did a mixture of the 2 interventions above along with analysing words using their phonemic elements (e.g. ‘at’ made up of ‘a’ and ‘t’) and using phonemes and rhymes in word building. <p>Control: ”children remained in class and undertook the word-level work appropriate to the second term of Year 1 of the National Literacy Strategy in their normal fashion“ (quote, p 55)</p> <p>Procedure: LSAs conducted training in small groups (typically 4 children per group - as per email from Savage on 30 November 2011). 20-minute sessions, 4 times/week, for a period of 9 weeks at school</p>
<p>Outcomes</p>	<p>Time of post-test: the week after training was completed</p> <p>Primary and secondary outcomes: letter-sound knowledge (experimental: ”cards with 26 individual letters on them“ (quote, p. 51) and phoneme awareness (experimental: nursery rhymes, rhyme matching, rhyme generation, blending and segmentation; see note 2 below)</p>
<p>Notes</p>	<ol style="list-style-type: none"> 1. Contacted Savage (on 24 January 2012) about what measured phonological awareness and letter sounds, and on 11 February 2012 about decoding and training group sizes. Replied that phonological awareness was measured by nursery rhymes, rhyme matching, rhyme generation, blending and segmentation; letter sounds was measured by 1 experimental test; and decoding skills was measured by nonsense word reading, word reading and spelling, and letter-sound knowledge. We asked for the individual scores for each of these tests however he only had combined scores. Finally, training groups typically had 4 children each 2. We used the combined score for phonological awareness in our analysis. 3. We did not use the decoding skills measure because it was a mixed of multiple skills that we used in this review as separate outcomes.

	<p>Study start and end dates: not reported</p> <p>Funding: financial support for the collaboration and execution of the project provided by the JJ Trust and the Helen Arkell Dyslexia Association. Financial support for the analysis and revision of the work provided by McGill University new researcher start-up fund no. 100810</p> <p>Declaration/potential conflicts of interest: none reported</p>	
<i>Risk of bias</i>		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	<p>Quote from publication: "a quasi-random allocation of schools to programs was undertaken: four schools whose catchment areas were known to draw primarily from lower SES backgrounds were each allocated to separate intervention groups. After that, for the other schools the allocation was entirely arbitrary... Children were, however, entirely arbitrarily allocated to an intervention condition (nine children) or to a control condition (three children)... As the allocation of children to intervention condition was not entirely arbitrary, but contained a systematic element..." (p 552)</p> <p>Quote from personal communication: "The same [as the Savage 2003 study] except that 4 schools of known low socio-economic status were each randomly allocated to one of the 4 groups first, using a random number generator. Then the process was repeated as above for all remaining schools. Child-level allocation was again undertaken using random number generator."</p>
Allocation concealment (selection bias)	Low risk	<p>Quote from personal communication: "I did this allocation independent of those running the study and of co-author(s) Carless and Stuart. Carless led the TA training, so I judge allocation to be concealed, and not possible to predict."</p> <p>Comment: could not foresee assignment due to central allocation of participants to groups</p>
Blinding of participants and personnel (performance bias) All outcomes	Low risk	<p>Quote from publication: "teachers were told who the control children and intervention children were, and were also reinforced</p>

		<p>at training and during the intervention to treat the control children in the same way as they would if no intervention was taking place for other children” (p 55)</p> <p>Quote from personal communication: “The TAs delivered [the training] based on sub-lexical phonological unit taught (rimes or phonemes) and this content is quite visible in the ‘treatment’ (no equivalent to a pill or placebo an option here). The one aspect that was blind was that we emphasized to TAs and all other school staff that each of the interventions (rime phoneme or mixed) was a proven evidence-based intervention, so we cast it as 3-horse race between them (with no favoured intervention) at all times, and emphasized the need for a ‘fair-test’ of each. TAs understood this. At the participant end, these are 6 years olds in both studies. They simply knew they were in an intervention (intervention condition children only of course) or receiving regular classroom teaching (control group children).”</p> <p>Comment: participants were children with little understanding of reading treatment techniques and hence were unlikely to understand allocation</p>
<p>Blinding of outcome assessment (detection bias) All outcomes</p>	<p>Low risk</p>	<p>Quote from personal communication: “Pre-testing was undertaken as a screen of all children in schools before we identified and allocated the ‘at-risk readers’, (see consort flow diagrams in both papers) so in this sense it is entirely blind... There was no blinding of post-testing in relation to the intervention condition as TAs did both (though see comments above on the 3 horse race). However classroom assistants also did not know of the theoretical contrasts TAs were not told at any point of any research predictions regarding the relationship between intervention and outcome (e.g. hypothesis of possible link between phoneme-based intervention and raise phoneme awareness at post-test, and similar for rimes etc.)”</p>

Savage 2005 (Continued)

Incomplete outcome data (attrition bias) All outcomes	Low risk	Quote from publication: “One child per intervention group was unavailable, having moved away from the LSA in the interim between pre- and post-test” (p 55) Comment: both groups experienced the same (relatively low) dropout rate
Selective reporting (reporting bias)	Unclear risk	Comment: data reported for all outcome measures outlined in methods; adequate detail for data to be included in analysis
Other bias	Low risk	Comment: none apparent

BAS: British Ability Scales; CC2: Castles and Coltheart 2; CSS: Classroom Survival Skills; DF: degrees-of-freedom; FSIQ: Full Scale IQ; GFW: Goldman-Fristoe-Woodcock; GPC: grapheme-to-phoneme correspondence; GV: garden variety; IQ: intelligence quotient; LSA: Learning Support Assistant; MinimPy: minimisation program; n: number of participants; PhAB: phonological analysis and blending; PI: principal investigator; RAN: rapid automatized naming; RD: reading difficulties; SD: standard deviation; SES: socioeconomic status; TA: teacher assistant; TOWRE: Test of Word Reading Efficiency; WIAT-II: Wechsler Individual Achievement Test Second Edition; WIST: Word Identification Strategy Training; WJRT: Woodcock-Johnson Reading Mastery Test; WJTA-III: Woodcock-Johnson Test of Achievement III; WRAT-R: Wide Range Achievement Test; WRMT-R: Woodcock Reading Mastery Test-Revised.

Characteristics of excluded studies [ordered by study ID]

Study	Reason for exclusion
Aboud 2018	Training did not match this review’s criteria for phonics training (Types of interventions); participants did not meet this review’s criteria for participants (Types of participants).
Alexander 1991	Trial did not include control data (Types of interventions).
Arnold 2016	Participants did not meet this review’s criteria for participants (Types of participants).
Berninger 2013	Training did not match this review’s criteria for phonics training (Types of interventions); trial did not include control data.
Bhide 2013	Participants did not meet this review’s criteria for participants (Types of participants).
Christodoulou 2017	Participants did not meet this review’s criteria for participants (Types of participants).
Dubois 2014	Participants did not meet this review’s criteria for participants (Types of participants).
Foorman 1997	Training did not match this review’s criteria for phonics training (Types of interventions).

(Continued)

Foorman 1998	Training did not match this review's criteria for phonics training (Types of interventions).
Gillon 1997	Training did not match this review's criteria for phonics training (Types of interventions).
Gillon 2000	Group allocation did not use randomisation, quasi-randomisation, or minimisation (Types of studies).
Gillon 2002	Group allocation did not use randomisation, quasi-randomisation, or minimisation (Types of studies).
Goldstein 2017	Training did not match this review's criteria for phonics training (Types of interventions).
Gorard 2015	Training did not match this review's criteria for phonics training (Types of interventions).
Hatcher 1994	Training did not match this review's criteria for phonics training (Types of interventions).
Hatcher 2006	Training did not match this review's criteria for phonics training (Types of interventions).
Jeffes 2016	Training did not match this review's criteria for phonics training (Types of interventions).
King 2015	Training did not match this review's criteria for phonics training (Types of interventions).
Lovett 1988	Training did not match this review's criteria for phonics training (Types of interventions).
Lovett 1989	Training did not match this review's criteria for phonics training (Types of interventions).
Lovett 1994	Participants did not meet this review's criteria for participants (Types of participants).
Lovett 2012	Training did not match this review's criteria for phonics training (Types of interventions).
Merrell 2015	Training did not match this review's criteria for phonics training (Types of interventions).
Messer 2018	Training did not match this review's criteria for phonics training (Types of interventions); participants did not meet this review's criteria for participants (Types of participants).
Metsala 2017	Training did not match this review's criteria for phonics training (Types of interventions).
Munro 2017	Training did not match this review's criteria for phonics training (Types of interventions).
Olson 1992	Review paper (Types of studies)
Olson 1997	Training did not match this review's criteria for phonics training (Types of interventions).
Rashotte 2001	Training did not match this review's criteria for phonics training (Types of interventions).
Savage 2018	Participants did not meet this review's criteria for participants (Types of participants).
Schaars 2017	Participants did not meet this review's criteria for participants (Types of participants).

(Continued)

Schlesinger 2017	Training did not match this review's criteria for phonics training (Types of interventions).
Seiler 2018	Training did not match this review's criteria for phonics training (Types of interventions).
Steady 2016	Training did not match this review's criteria for phonics training (Types of interventions).
Storey 2017	Training did not match this review's criteria for phonics training (Types of interventions).
Torgesen 1997	Reading was not assessed pre- and post-training.
Torgesen 1999a	Training did not match this review's criteria for phonics training (Types of interventions).
Torgesen 2001	Training did not match this review's criteria for phonics training (Types of interventions); trial did not include control data.
Torgesen 2006	Training did not match this review's criteria for phonics training (Types of interventions).
Van Gorp 2017	Participants did not meet this review's criteria for participants (Types of participants).
Vellutino 1986	Training did not match this review's criteria for phonics training (Types of interventions).
Vellutino 1987	Training did not match this review's criteria for phonics training (Types of interventions).
Vellutino 1996	Training did not match this review's criteria for phonics training (Types of interventions).
Wheldall 2017	Training did not match this review's criteria for phonics training (Types of interventions).
Wise 1995	Training did not match this review's criteria for phonics training (Types of interventions); trial did not include control data.
Wise 1997	Training did not match this review's criteria for phonics training (Types of interventions).
Wise 1999	Training did not match this review's criteria for phonics training (Types of interventions).
Wise 2000	Training did not match this review's criteria for phonics training (Types of interventions); trial did not include control data.

DATA AND ANALYSES

Comparison 1. Phonics training versus control (random-effects model)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Mixed/regular word reading accuracy	11	701	Std. Mean Difference (IV, Random, 95% CI)	0.51 [0.13, 0.90]
2 Non-word reading accuracy	10	682	Std. Mean Difference (IV, Random, 95% CI)	0.67 [0.26, 1.07]
3 Irregular word reading accuracy	4	294	Std. Mean Difference (IV, Random, 95% CI)	0.84 [0.30, 1.39]
4 Mixed/regular word reading fluency	4	224	Std. Mean Difference (IV, Random, 95% CI)	0.45 [0.19, 0.72]
5 Non-word reading fluency	3	188	Std. Mean Difference (IV, Random, 95% CI)	0.39 [0.10, 0.68]
6 Reading comprehension	5	343	Std. Mean Difference (IV, Random, 95% CI)	0.28 [-0.07, 0.62]
7 Spelling	3	158	Std. Mean Difference (IV, Random, 95% CI)	0.47 [-0.07, 1.01]
8 Letter-sound knowledge	3	192	Std. Mean Difference (IV, Random, 95% CI)	0.35 [0.04, 0.65]
9 Phonological output	4	280	Std. Mean Difference (IV, Random, 95% CI)	0.38 [-0.04, 0.80]

Comparison 2. Phonics training versus control: subgroup analyses (random-effects model)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Mixed/regular word reading accuracy	11		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
1.1 Training type: phonics alone	3	232	Std. Mean Difference (IV, Random, 95% CI)	0.94 [-0.09, 1.97]
1.2 Training type: phonics + phoneme awareness	6	415	Std. Mean Difference (IV, Random, 95% CI)	0.17 [-0.04, 0.37]
1.3 Training type: phonics + sight words	2	54	Std. Mean Difference (IV, Random, 95% CI)	0.73 [0.18, 1.29]
1.4 Training intensity: < 2 hours/week	9	577	Std. Mean Difference (IV, Random, 95% CI)	0.54 [0.06, 1.02]
1.5 Training intensity: ≥ 2 hours/week	2	124	Std. Mean Difference (IV, Random, 95% CI)	0.34 [-0.02, 0.70]
1.6 Training duration: < 3 months	9	516	Std. Mean Difference (IV, Random, 95% CI)	0.61 [0.17, 1.05]
1.7 Training duration: ≥ 3 months	2	185	Std. Mean Difference (IV, Random, 95% CI)	0.12 [-0.43, 0.67]
1.8 Training group size: 1-on-1	6	419	Std. Mean Difference (IV, Random, 95% CI)	0.62 [-0.06, 1.29]
1.9 Training group size: small group (≤ 5)	5	282	Std. Mean Difference (IV, Random, 95% CI)	0.33 [0.04, 0.61]
1.10 Training administrator: human	7	577	Std. Mean Difference (IV, Random, 95% CI)	0.70 [0.17, 1.23]

1.11 Training administrator: computer	4	124	Std. Mean Difference (IV, Random, 95% CI)	0.18 [-0.17, 0.54]
2 Non-word reading accuracy	10		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
2.1 Training type: phonics alone	5	402	Std. Mean Difference (IV, Random, 95% CI)	0.69 [-0.08, 1.46]
2.2 Training type: phonics + phoneme awareness	5	280	Std. Mean Difference (IV, Random, 95% CI)	0.63 [0.38, 0.88]
2.3 Training group size: 1-on-1	7	454	Std. Mean Difference (IV, Random, 95% CI)	0.83 [0.31, 1.36]
2.4 Training group size: small group (≤ 5)	3	228	Std. Mean Difference (IV, Random, 95% CI)	0.32 [-0.32, 0.96]
2.5 Training administrator: human	4	388	Std. Mean Difference (IV, Random, 95% CI)	1.12 [0.48, 1.76]
2.6 Training administrator: computer	6	294	Std. Mean Difference (IV, Random, 95% CI)	0.31 [-0.02, 0.64]

Comparison 3. Phonics training versus control: sensitivity analysis using the fixed-effect model

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Mixed/regular word reading accuracy	11	701	Std. Mean Difference (IV, Fixed, 95% CI)	0.48 [0.32, 0.64]
2 Non-word reading accuracy	10	682	Std. Mean Difference (IV, Fixed, 95% CI)	0.68 [0.51, 0.84]
3 Irregular word reading accuracy	4	294	Std. Mean Difference (IV, Fixed, 95% CI)	0.82 [0.58, 1.07]
4 Mixed/regular word reading fluency	4	224	Std. Mean Difference (IV, Fixed, 95% CI)	0.45 [0.19, 0.72]
5 Non-word reading fluency	3	188	Std. Mean Difference (IV, Fixed, 95% CI)	0.39 [0.10, 0.68]
6 Reading comprehension	5	343	Std. Mean Difference (IV, Fixed, 95% CI)	0.23 [0.01, 0.45]
7 Spelling	2	140	Std. Mean Difference (IV, Fixed, 95% CI)	0.28 [-0.09, 0.65]
8 Letter-sound knowledge	3	192	Std. Mean Difference (IV, Fixed, 95% CI)	0.35 [0.04, 0.65]
9 Phonological output	4	280	Std. Mean Difference (IV, Fixed, 95% CI)	0.44 [0.19, 0.70]

Comparison 4. Phonics training versus control: sensitivity analysis with Hurford 1994 removed (random-effects model)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Mixed/regular word reading accuracy	10	651	Std. Mean Difference (IV, Random, 95% CI)	0.52 [0.09, 0.95]
2 Non-word reading accuracy	9	632	Std. Mean Difference (IV, Random, 95% CI)	0.69 [0.24, 1.14]

Comparison 5. Phonics training versus control: sensitivity analysis with small studies removed (n < 11)

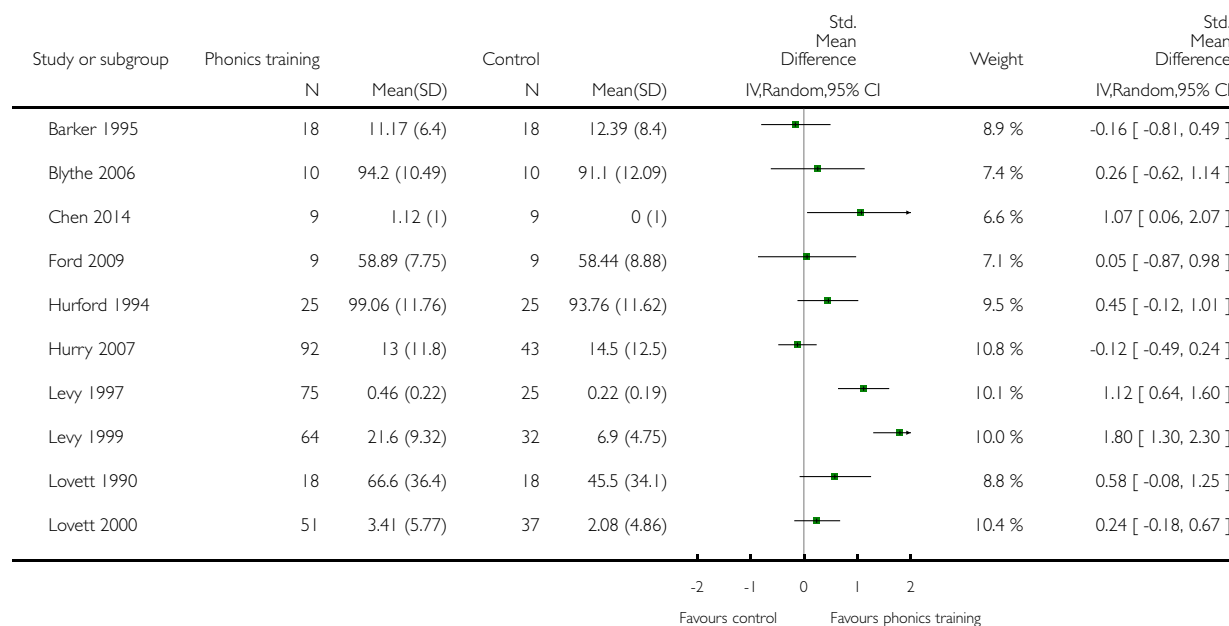
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Mixed/regular word reading accuracy	8	645	Std. Mean Difference (IV, Random, 95% CI)	0.53 [0.07, 1.00]
2 Non-word reading accuracy	8	644	Std. Mean Difference (IV, Random, 95% CI)	0.66 [0.20, 1.11]
3 Irregular word reading accuracy	4	294	Std. Mean Difference (IV, Random, 95% CI)	0.84 [0.30, 1.39]
4 Mixed/regular word reading fluency	3	206	Std. Mean Difference (IV, Random, 95% CI)	0.50 [0.22, 0.78]
5 Non-word reading fluency	2	170	Std. Mean Difference (IV, Random, 95% CI)	0.39 [0.08, 0.69]
6 Reading comprehension	3	305	Std. Mean Difference (IV, Random, 95% CI)	0.25 [-0.15, 0.64]
7 Spelling	2	140	Std. Mean Difference (IV, Random, 95% CI)	0.36 [-0.27, 0.99]
8 Letter-sound knowledge	3	192	Std. Mean Difference (IV, Random, 95% CI)	0.35 [0.04, 0.65]
9 Phonological output	4	280	Std. Mean Difference (IV, Random, 95% CI)	0.38 [-0.04, 0.80]

Analysis 1.1. Comparison 1 Phonics training versus control (random-effects model), Outcome 1 Mixed/regular word reading accuracy.

Review: Phonics training for English-speaking poor readers

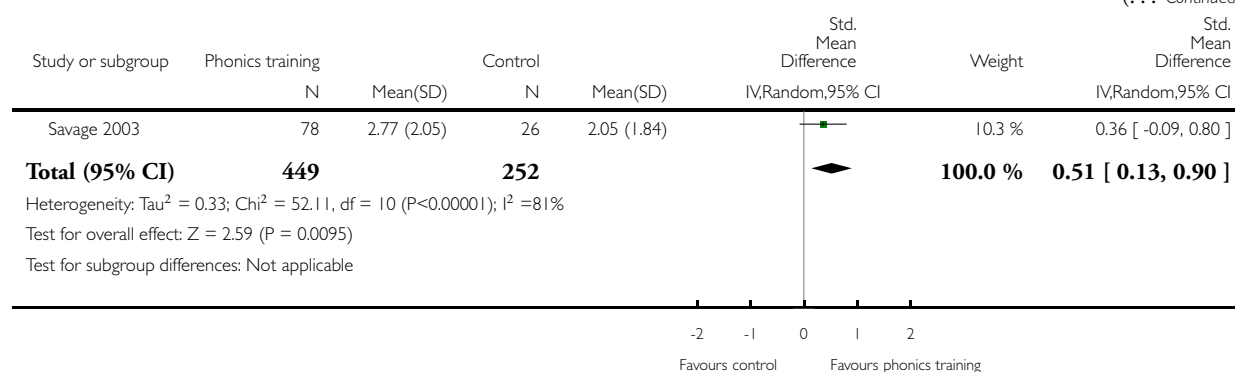
Comparison: 1 Phonics training versus control (random-effects model)

Outcome: 1 Mixed/regular word reading accuracy



(Continued . . .)

(... Continued)

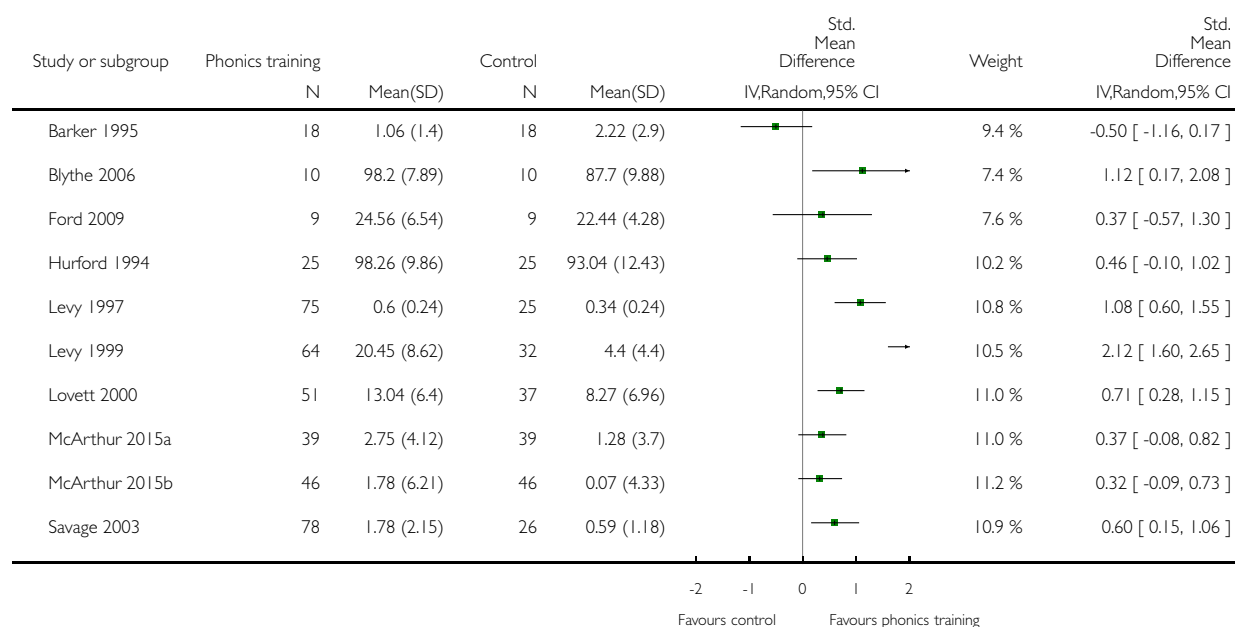


Analysis 1.2. Comparison 1 Phonics training versus control (random-effects model), Outcome 2 Non-word reading accuracy.

Review: Phonics training for English-speaking poor readers

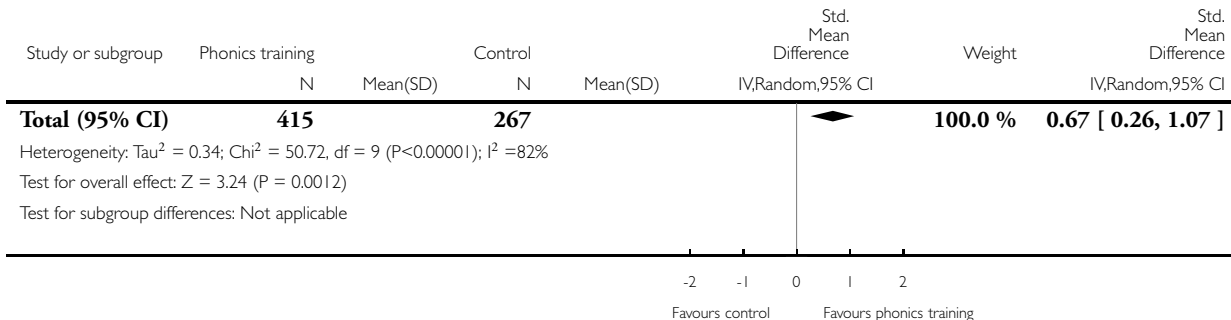
Comparison: 1 Phonics training versus control (random-effects model)

Outcome: 2 Non-word reading accuracy



(Continued ...)

(... Continued)

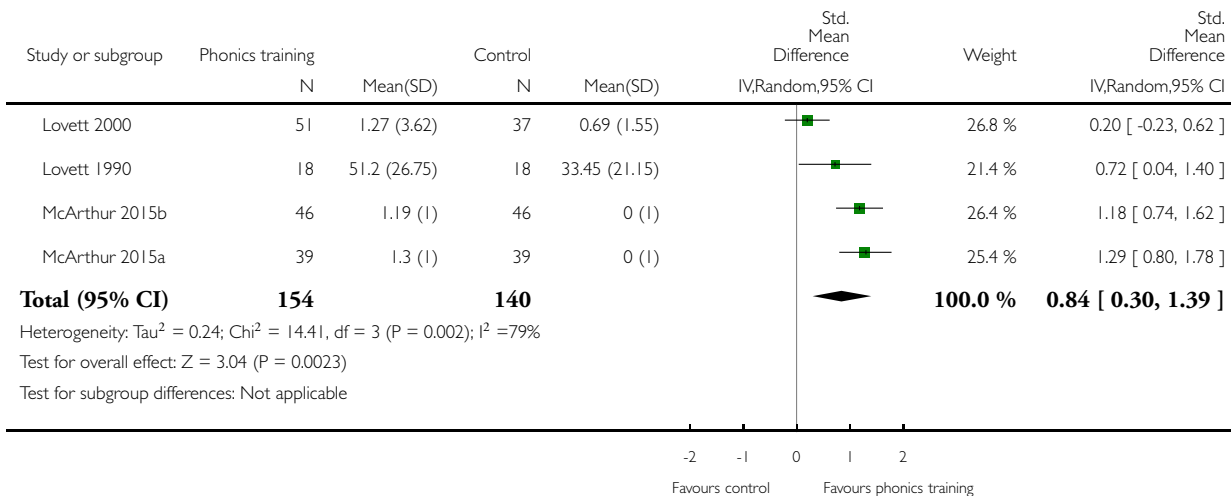


Analysis 1.3. Comparison 1 Phonics training versus control (random-effects model), Outcome 3 Irregular word reading accuracy.

Review: Phonics training for English-speaking poor readers

Comparison: 1 Phonics training versus control (random-effects model)

Outcome: 3 Irregular word reading accuracy

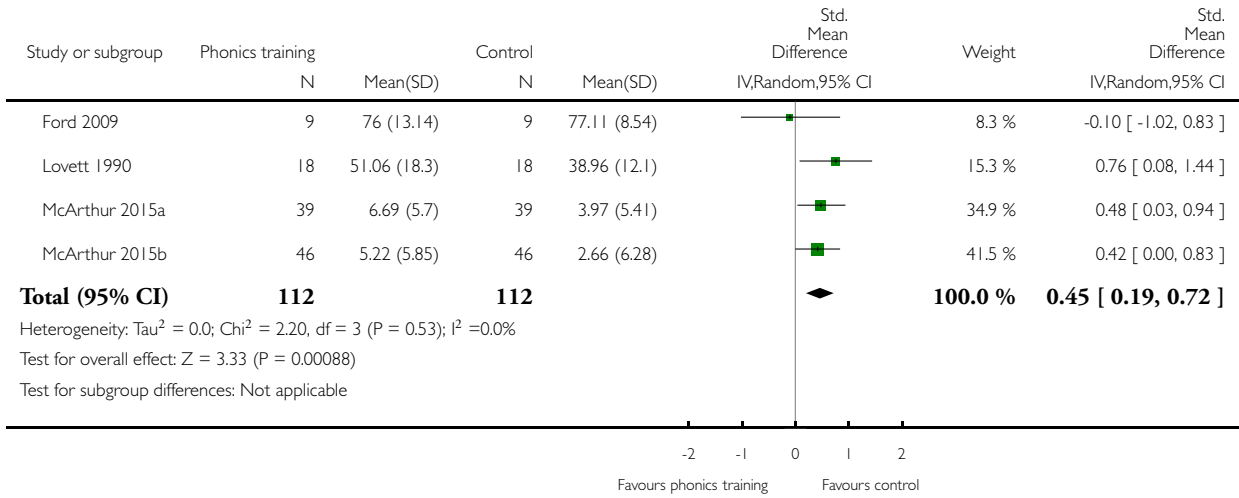


Analysis 1.4. Comparison 1 Phonics training versus control (random-effects model), Outcome 4 Mixed/regular word reading fluency.

Review: Phonics training for English-speaking poor readers

Comparison: 1 Phonics training versus control (random-effects model)

Outcome: 4 Mixed/regular word reading fluency

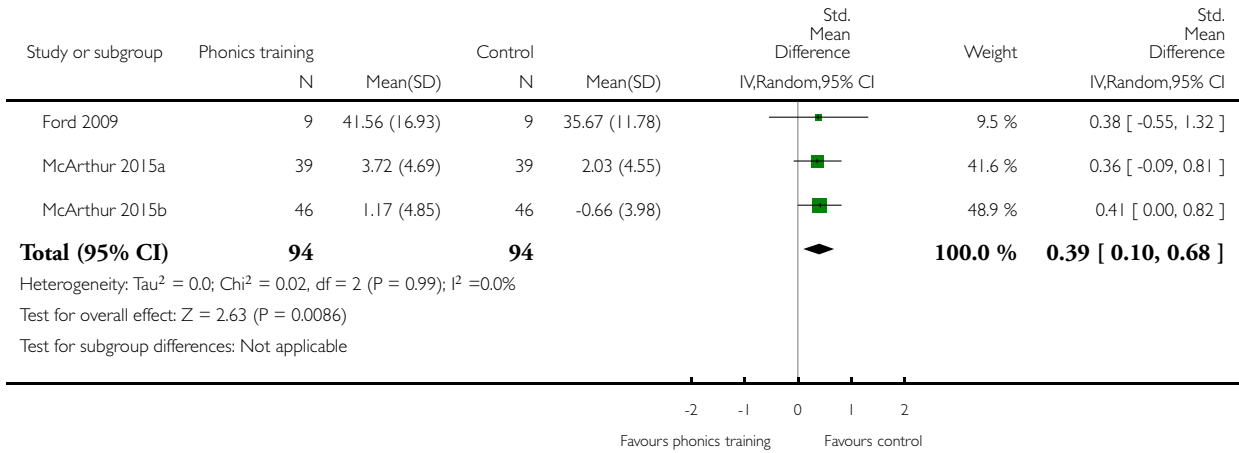


Analysis 1.5. Comparison 1 Phonics training versus control (random-effects model), Outcome 5 Non-word reading fluency.

Review: Phonics training for English-speaking poor readers

Comparison: 1 Phonics training versus control (random-effects model)

Outcome: 5 Non-word reading fluency

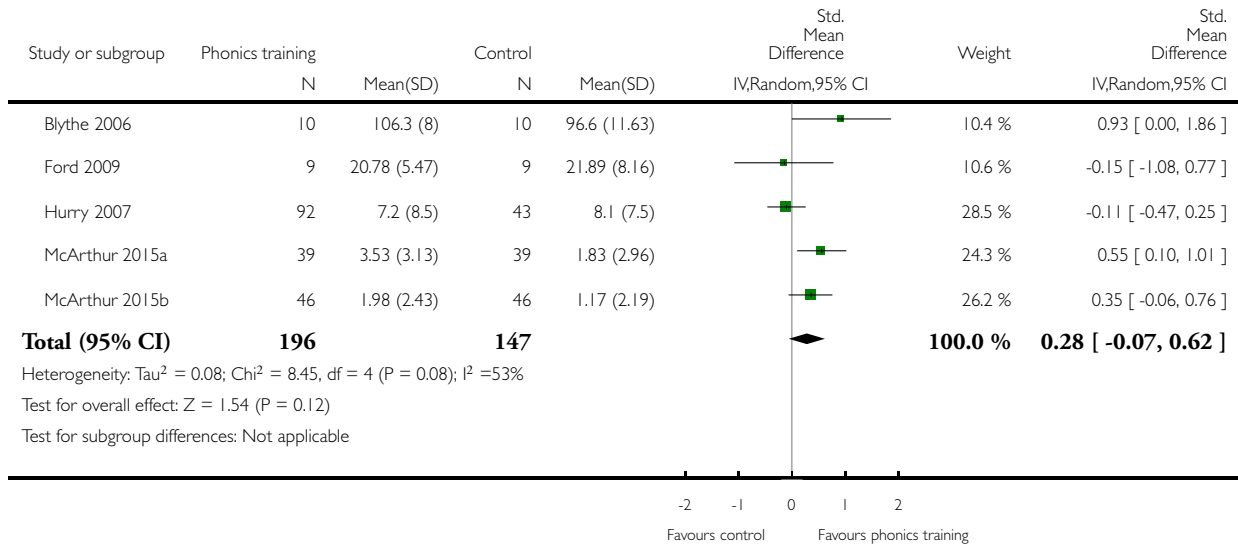


Analysis 1.6. Comparison 1 Phonics training versus control (random-effects model), Outcome 6 Reading comprehension.

Review: Phonics training for English-speaking poor readers

Comparison: 1 Phonics training versus control (random-effects model)

Outcome: 6 Reading comprehension

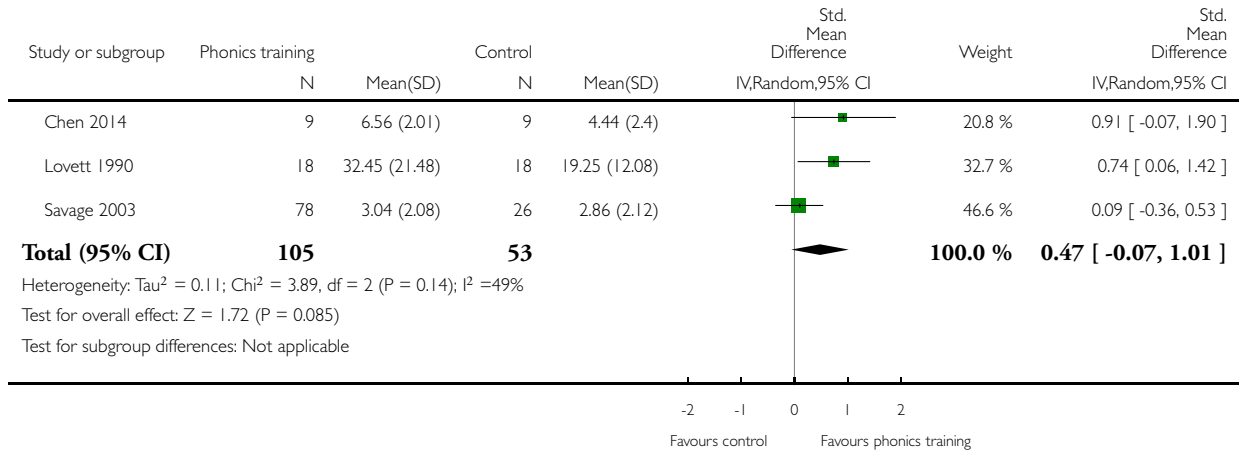


Analysis 1.7. Comparison 1 Phonics training versus control (random-effects model), Outcome 7 Spelling.

Review: Phonics training for English-speaking poor readers

Comparison: 1 Phonics training versus control (random-effects model)

Outcome: 7 Spelling

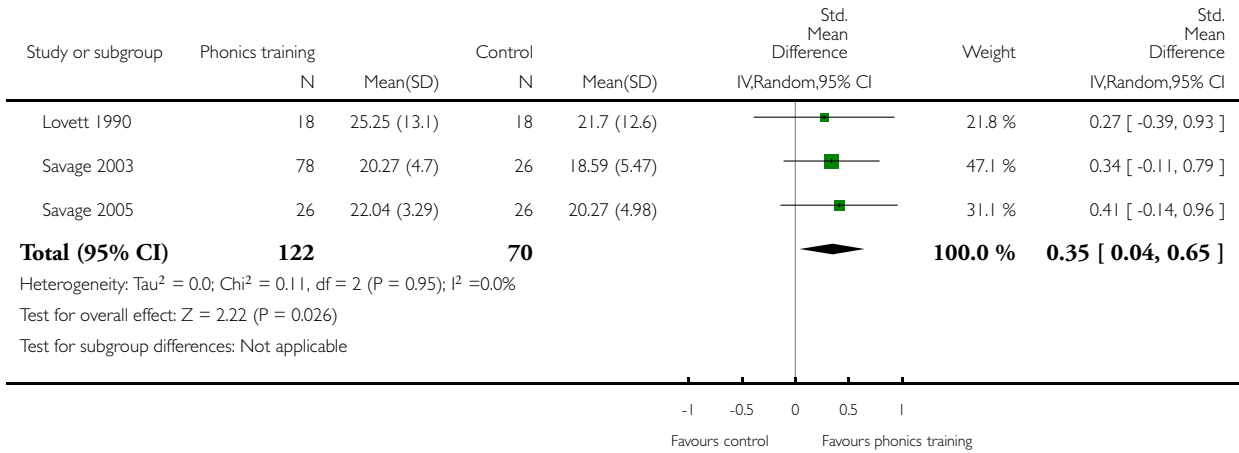


Analysis 1.8. Comparison 1 Phonics training versus control (random-effects model), Outcome 8 Letter-sound knowledge.

Review: Phonics training for English-speaking poor readers

Comparison: 1 Phonics training versus control (random-effects model)

Outcome: 8 Letter-sound knowledge

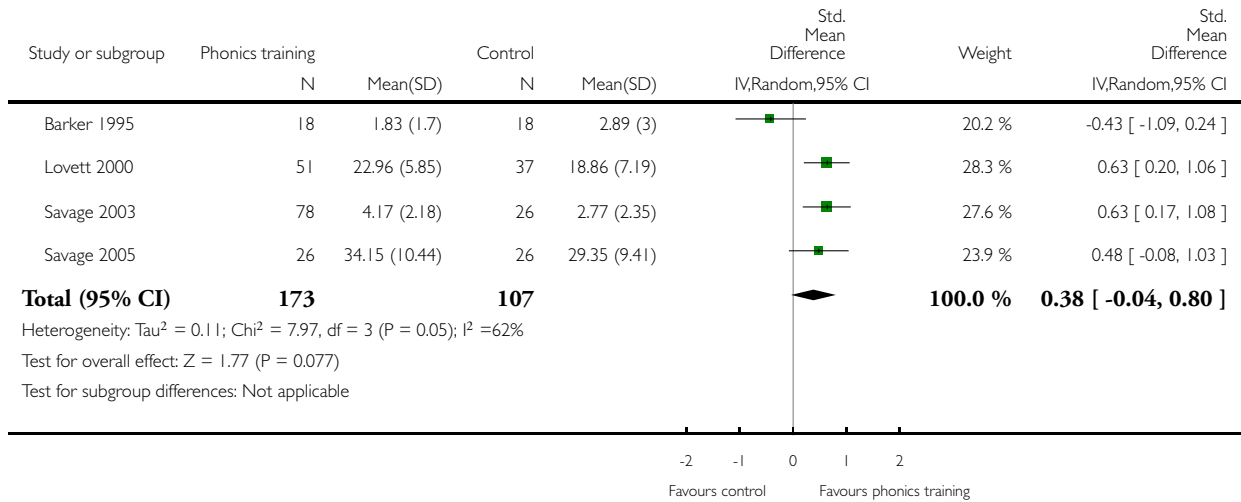


Analysis 1.9. Comparison 1 Phonics training versus control (random-effects model), Outcome 9 Phonological output.

Review: Phonics training for English-speaking poor readers

Comparison: 1 Phonics training versus control (random-effects model)

Outcome: 9 Phonological output

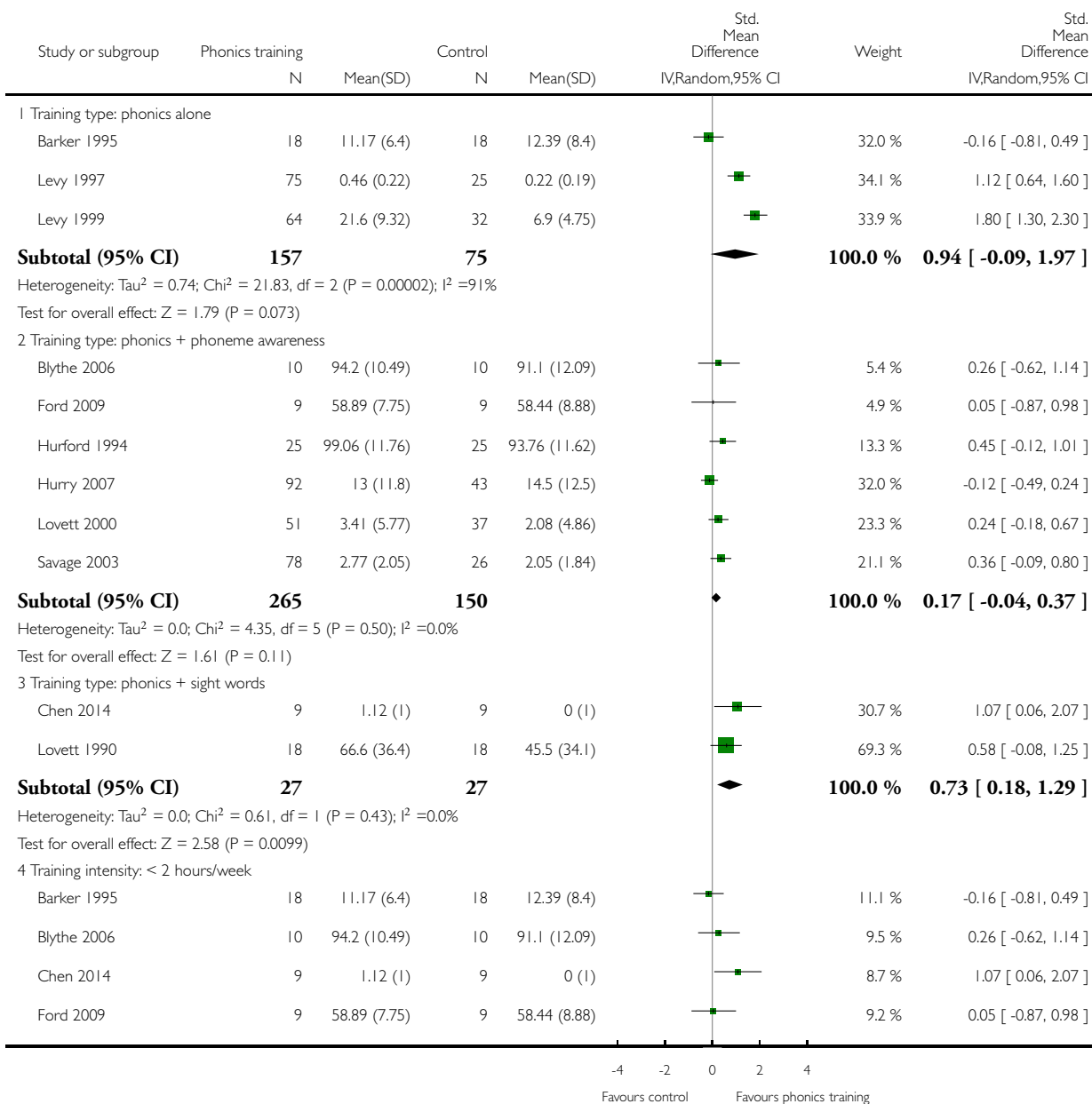


Analysis 2.1. Comparison 2 Phonics training versus control: subgroup analyses (random-effects model), Outcome 1 Mixed/regular word reading accuracy.

Review: Phonics training for English-speaking poor readers

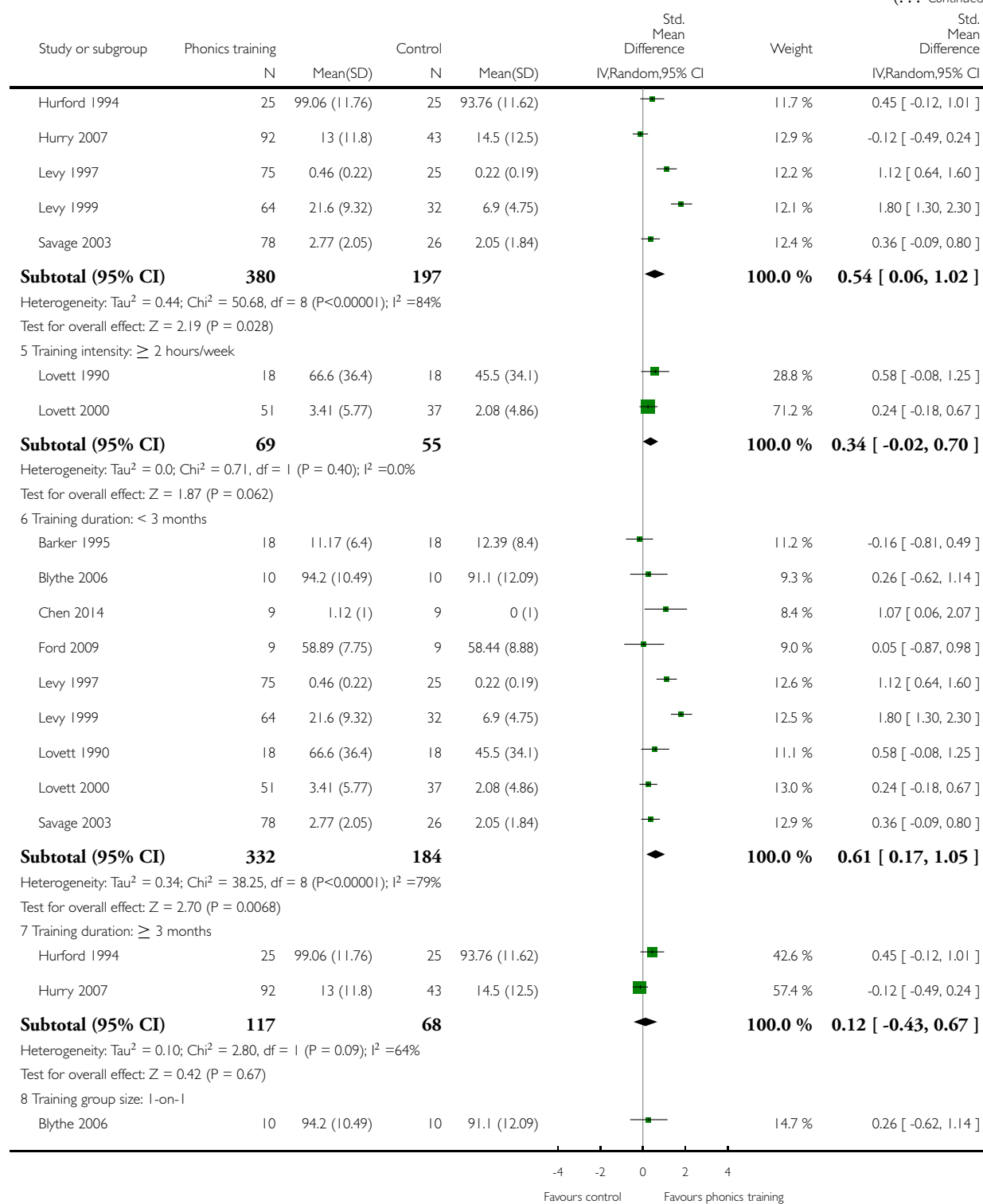
Comparison: 2 Phonics training versus control: subgroup analyses (random-effects model)

Outcome: 1 Mixed/regular word reading accuracy



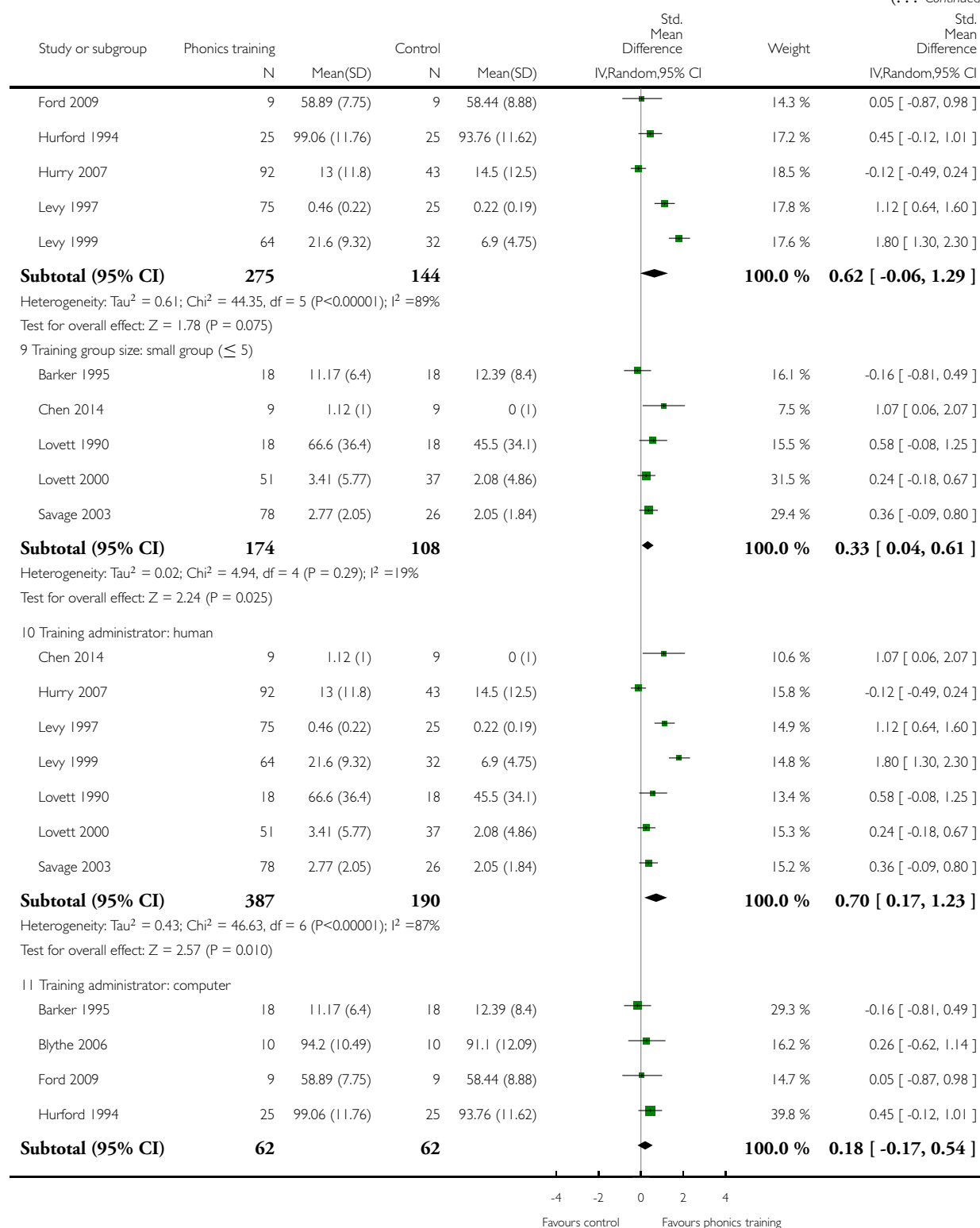
(Continued ...)

(... Continued)



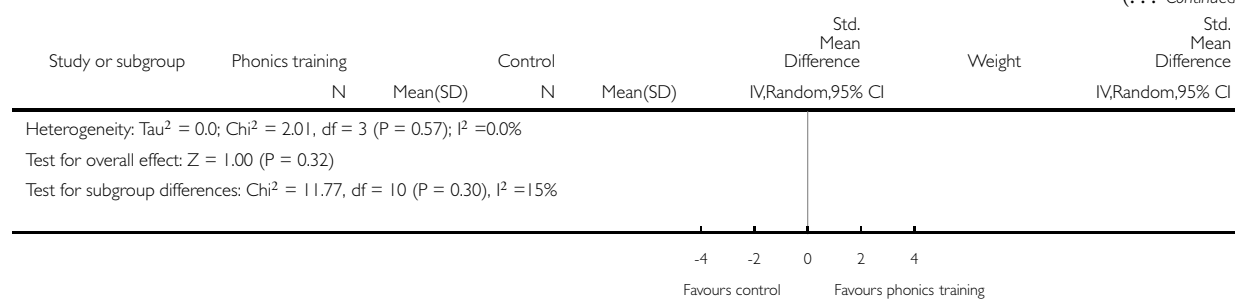
(Continued ...)

(... Continued)



(Continued ...)

(... Continued)

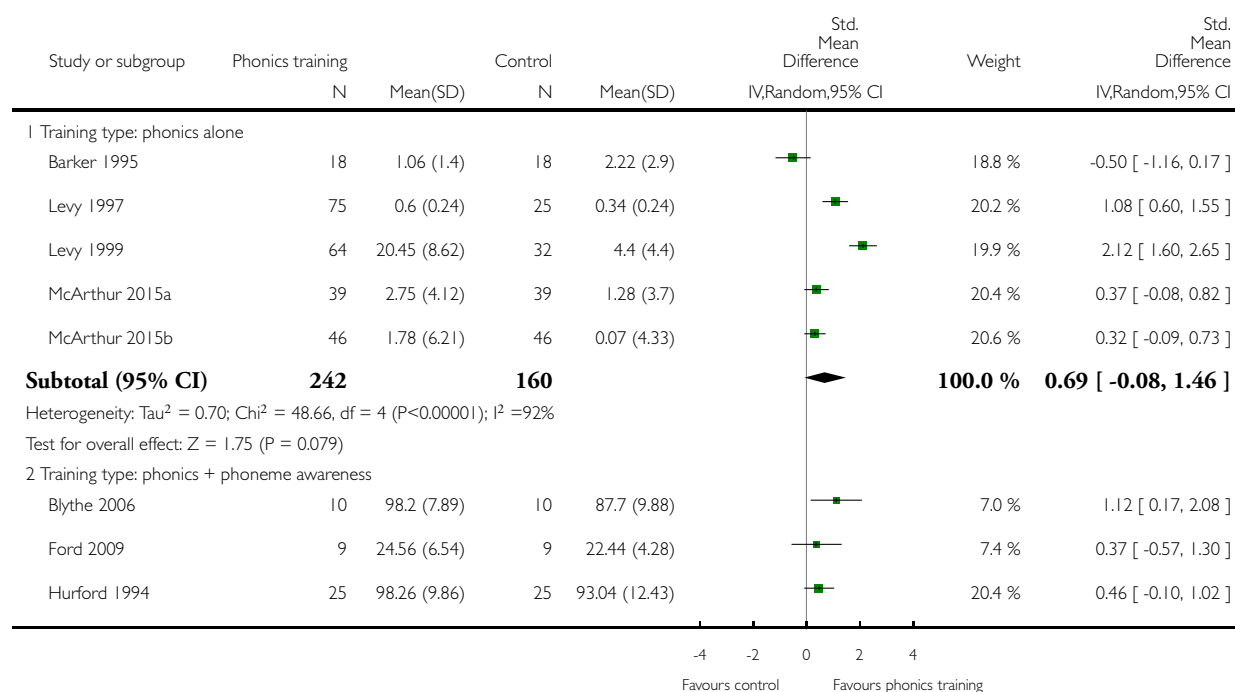


Analysis 2.2. Comparison 2 Phonics training versus control: subgroup analyses (random-effects model), Outcome 2 Non-word reading accuracy.

Review: Phonics training for English-speaking poor readers

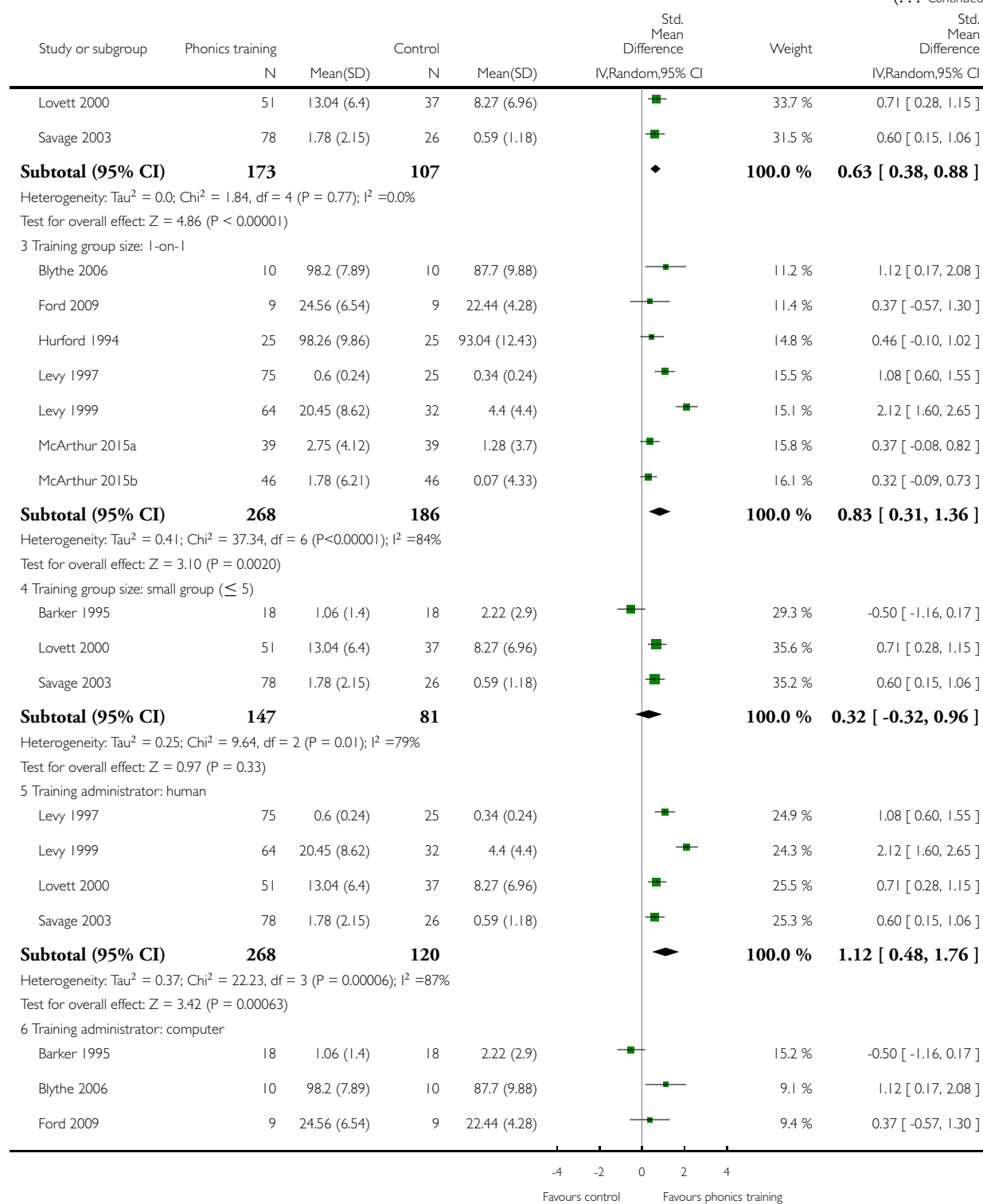
Comparison: 2 Phonics training versus control: subgroup analyses (random-effects model)

Outcome: 2 Non-word reading accuracy



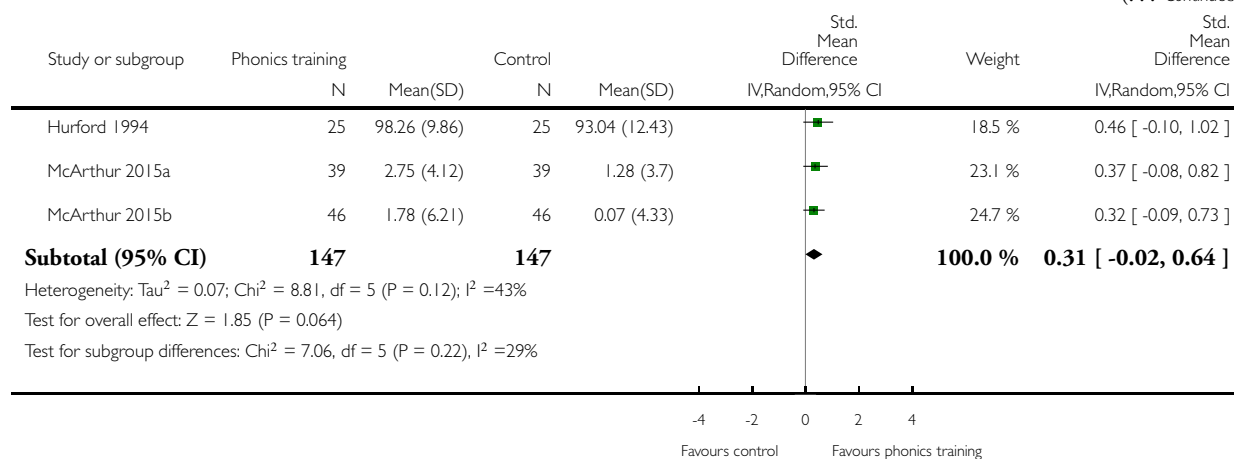
(Continued ...)

(... Continued)



(Continued ...)

(... Continued)

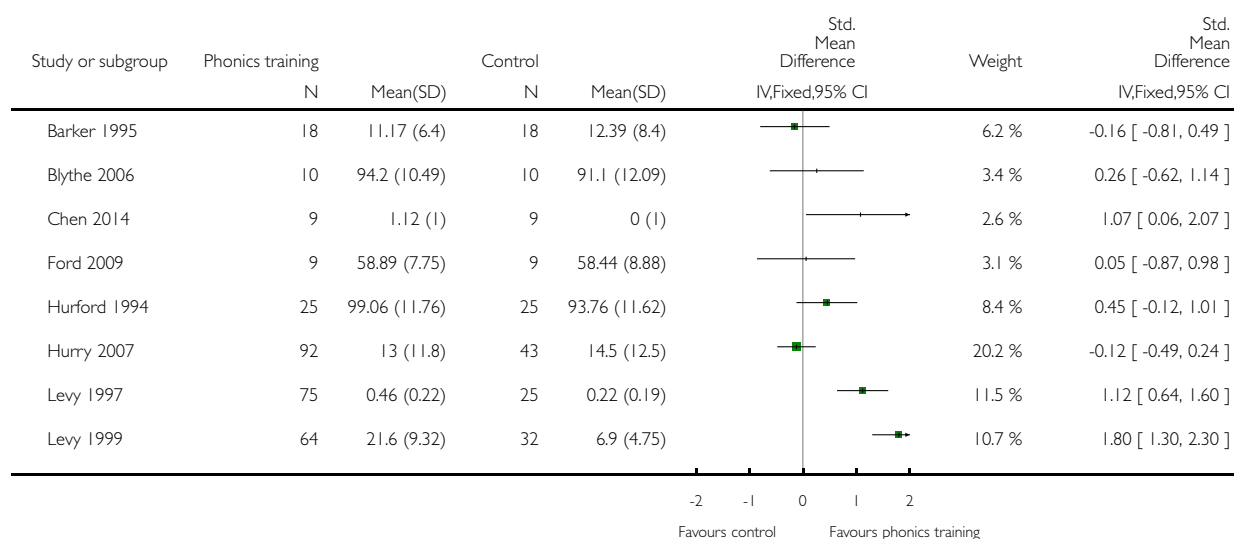


Analysis 3.1. Comparison 3 Phonics training versus control: sensitivity analysis using the fixed-effect model, Outcome 1 Mixed/regular word reading accuracy.

Review: Phonics training for English-speaking poor readers

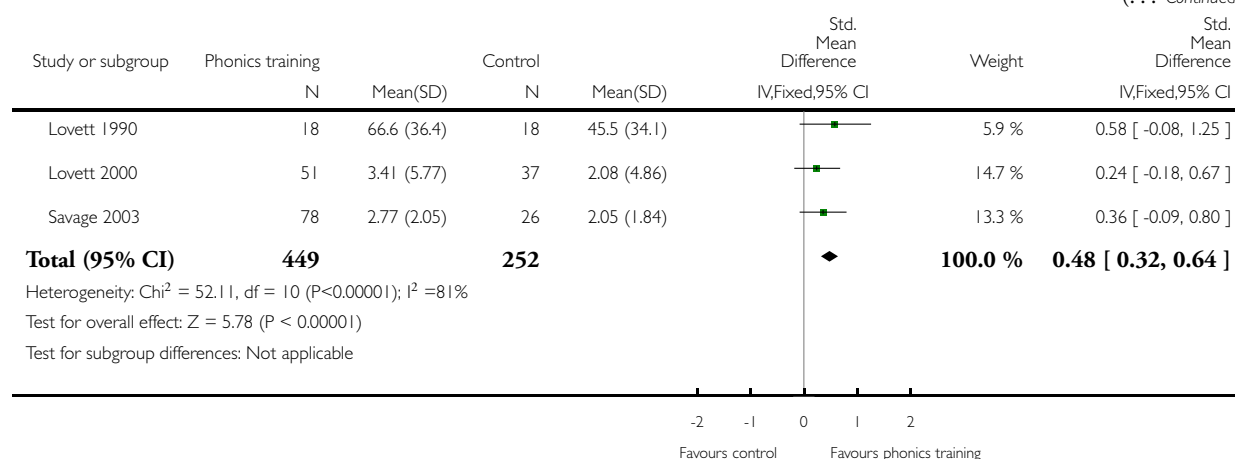
Comparison: 3 Phonics training versus control: sensitivity analysis using the fixed-effect model

Outcome: 1 Mixed/regular word reading accuracy



(Continued ...)

(... Continued)

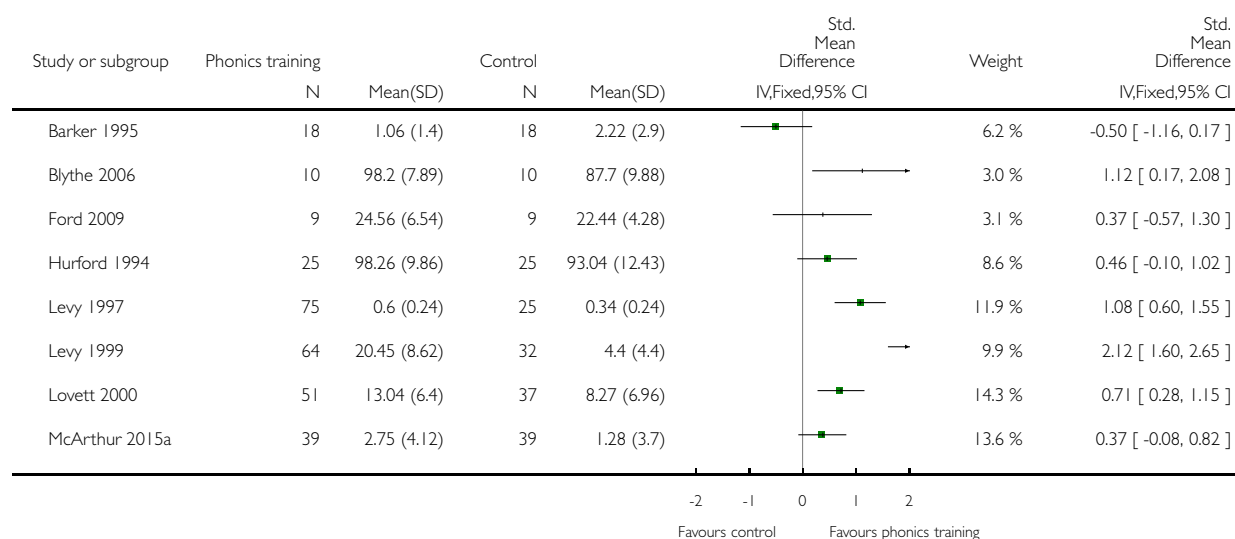


Analysis 3.2. Comparison 3 Phonics training versus control: sensitivity analysis using the fixed-effect model, Outcome 2 Non-word reading accuracy.

Review: Phonics training for English-speaking poor readers

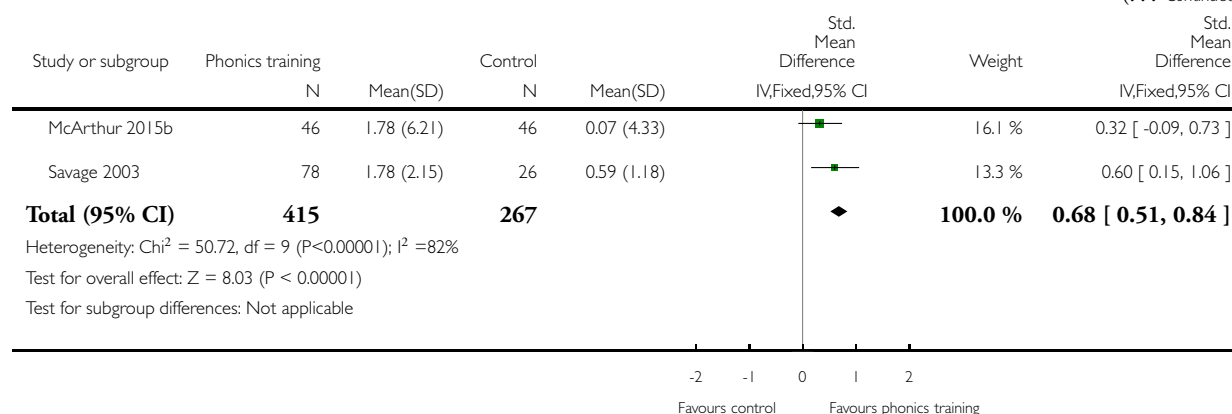
Comparison: 3 Phonics training versus control: sensitivity analysis using the fixed-effect model

Outcome: 2 Non-word reading accuracy



(Continued ...)

(... Continued)

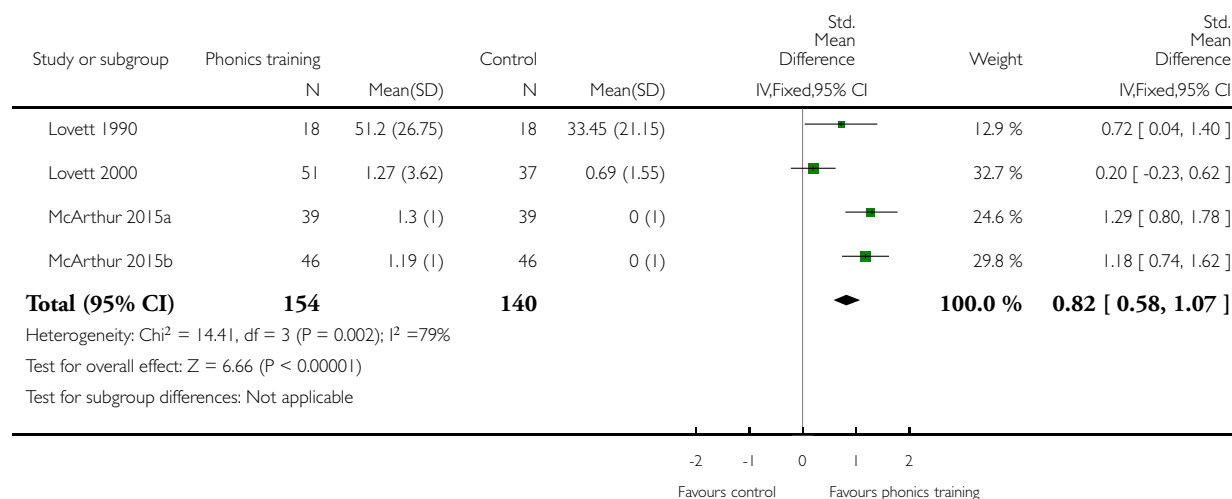


Analysis 3.3. Comparison 3 Phonics training versus control: sensitivity analysis using the fixed-effect model, Outcome 3 Irregular word reading accuracy.

Review: Phonics training for English-speaking poor readers

Comparison: 3 Phonics training versus control: sensitivity analysis using the fixed-effect model

Outcome: 3 Irregular word reading accuracy

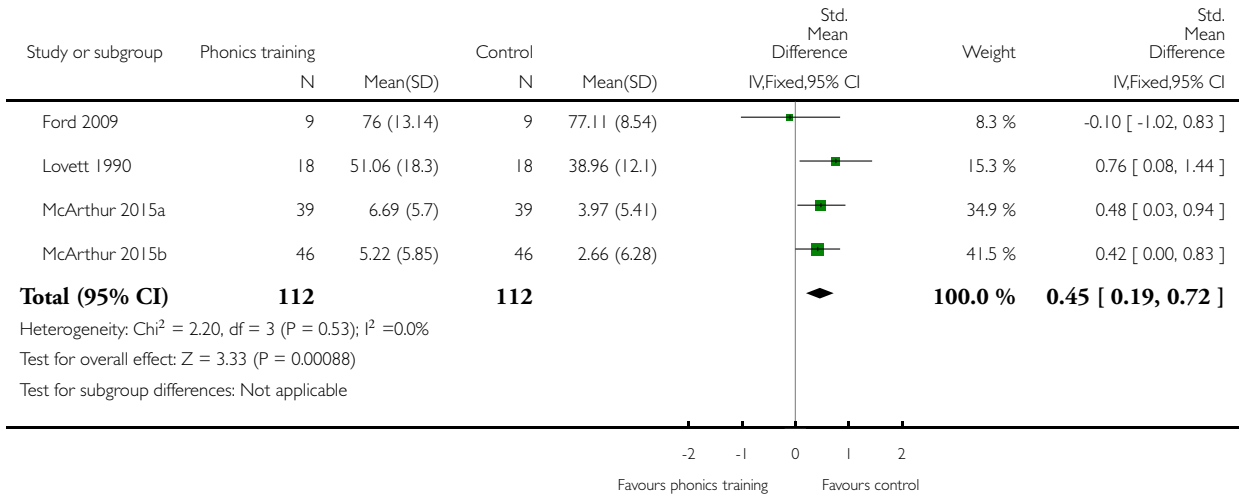


Analysis 3.4. Comparison 3 Phonics training versus control: sensitivity analysis using the fixed-effect model, Outcome 4 Mixed/regular word reading fluency.

Review: Phonics training for English-speaking poor readers

Comparison: 3 Phonics training versus control: sensitivity analysis using the fixed-effect model

Outcome: 4 Mixed/regular word reading fluency

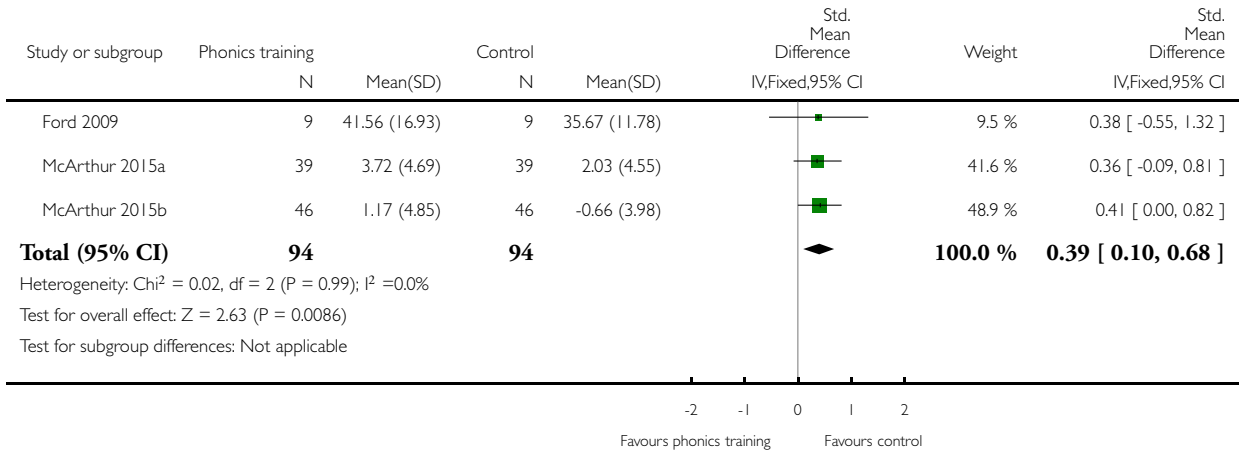


Analysis 3.5. Comparison 3 Phonics training versus control: sensitivity analysis using the fixed-effect model, Outcome 5 Non-word reading fluency.

Review: Phonics training for English-speaking poor readers

Comparison: 3 Phonics training versus control: sensitivity analysis using the fixed-effect model

Outcome: 5 Non-word reading fluency

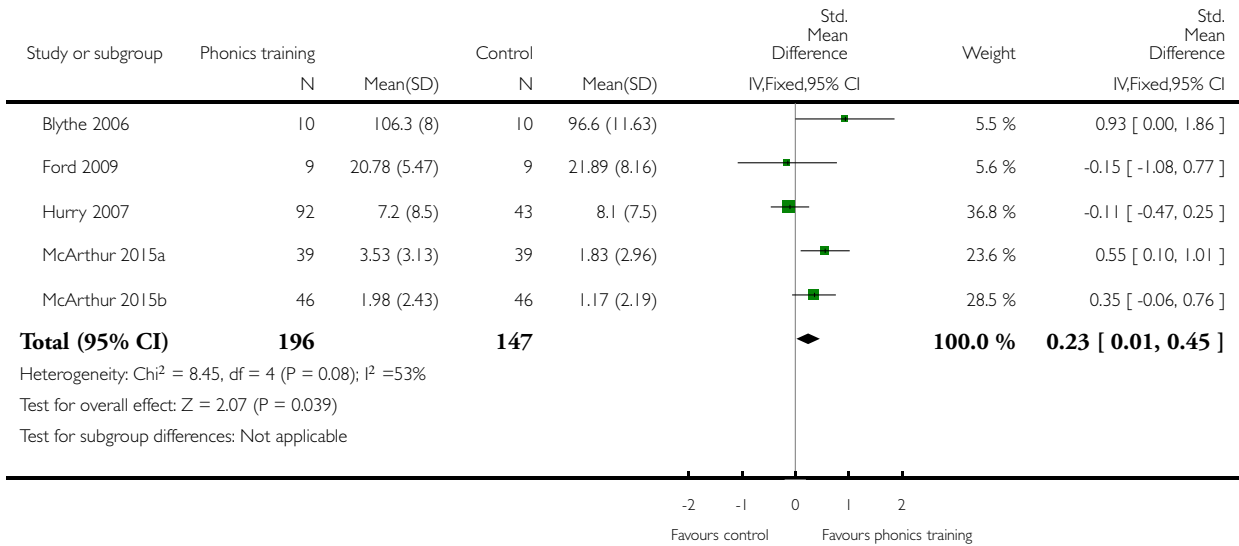


Analysis 3.6. Comparison 3 Phonics training versus control: sensitivity analysis using the fixed-effect model, Outcome 6 Reading comprehension.

Review: Phonics training for English-speaking poor readers

Comparison: 3 Phonics training versus control: sensitivity analysis using the fixed-effect model

Outcome: 6 Reading comprehension

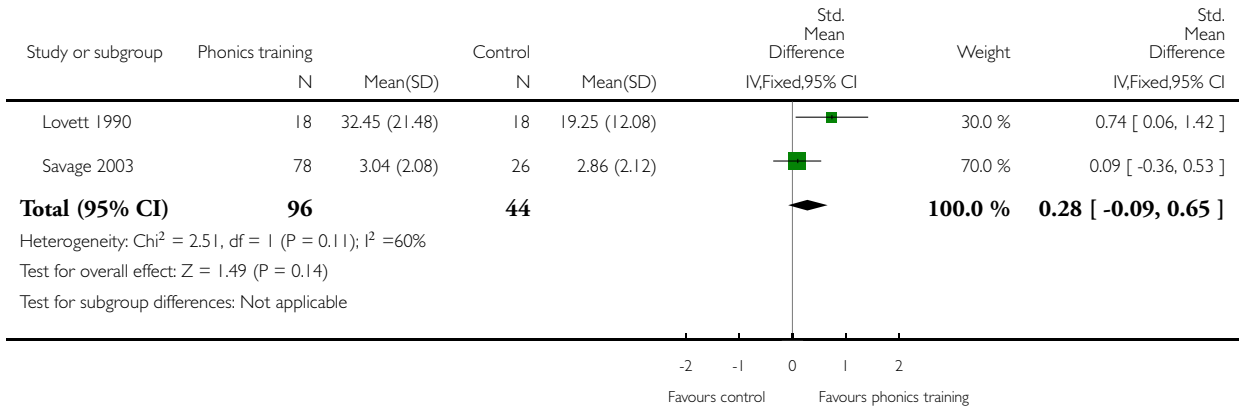


Analysis 3.7. Comparison 3 Phonics training versus control: sensitivity analysis using the fixed-effect model, Outcome 7 Spelling.

Review: Phonics training for English-speaking poor readers

Comparison: 3 Phonics training versus control: sensitivity analysis using the fixed-effect model

Outcome: 7 Spelling

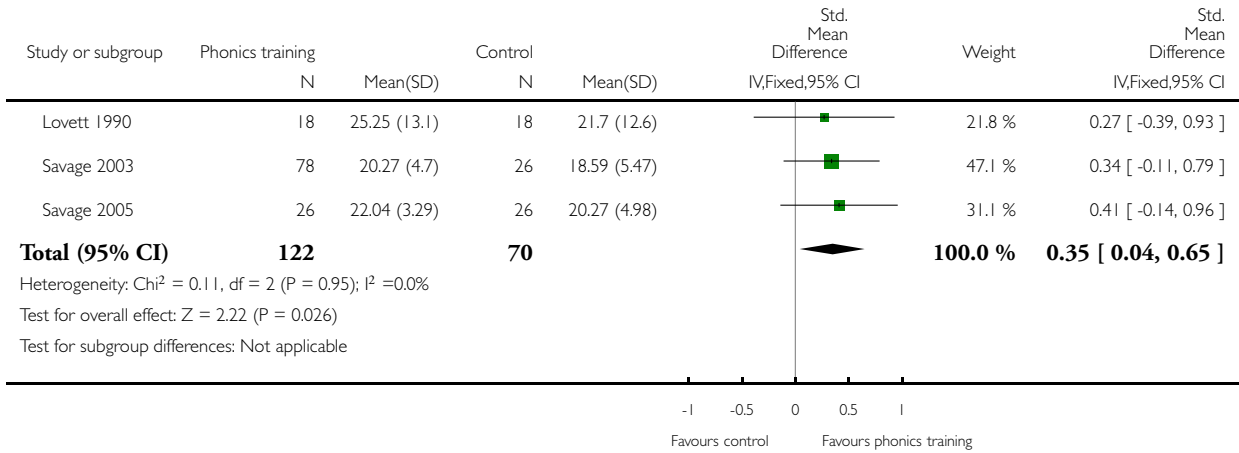


Analysis 3.8. Comparison 3 Phonics training versus control: sensitivity analysis using the fixed-effect model, Outcome 8 Letter-sound knowledge.

Review: Phonics training for English-speaking poor readers

Comparison: 3 Phonics training versus control: sensitivity analysis using the fixed-effect model

Outcome: 8 Letter-sound knowledge

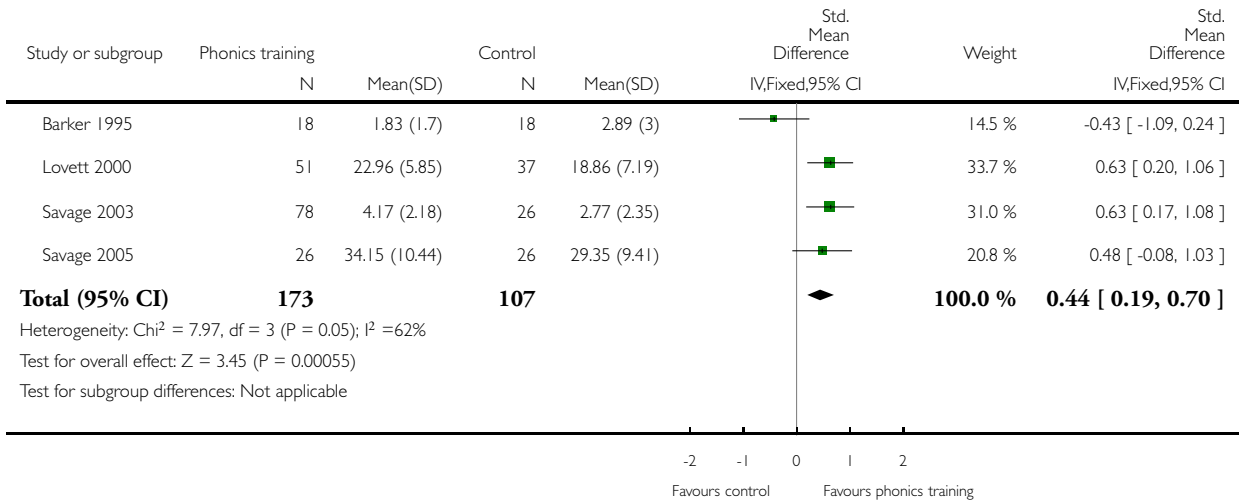


Analysis 3.9. Comparison 3 Phonics training versus control: sensitivity analysis using the fixed-effect model, Outcome 9 Phonological output.

Review: Phonics training for English-speaking poor readers

Comparison: 3 Phonics training versus control: sensitivity analysis using the fixed-effect model

Outcome: 9 Phonological output

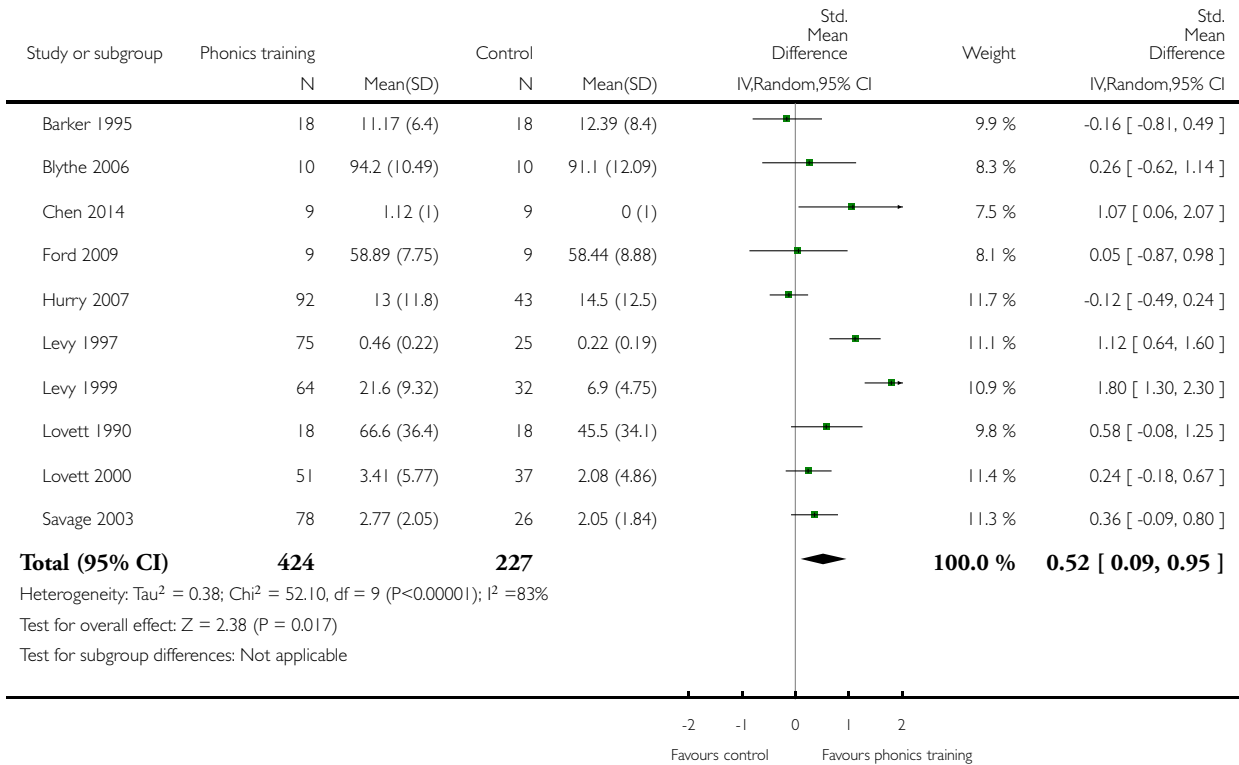


Analysis 4.1. Comparison 4 Phonics training versus control: sensitivity analysis with Hurford 1994 removed (random-effects model), Outcome 1 Mixed/regular word reading accuracy.

Review: Phonics training for English-speaking poor readers

Comparison: 4 Phonics training versus control: sensitivity analysis with Hurford 1994 removed (random-effects model)

Outcome: 1 Mixed/regular word reading accuracy

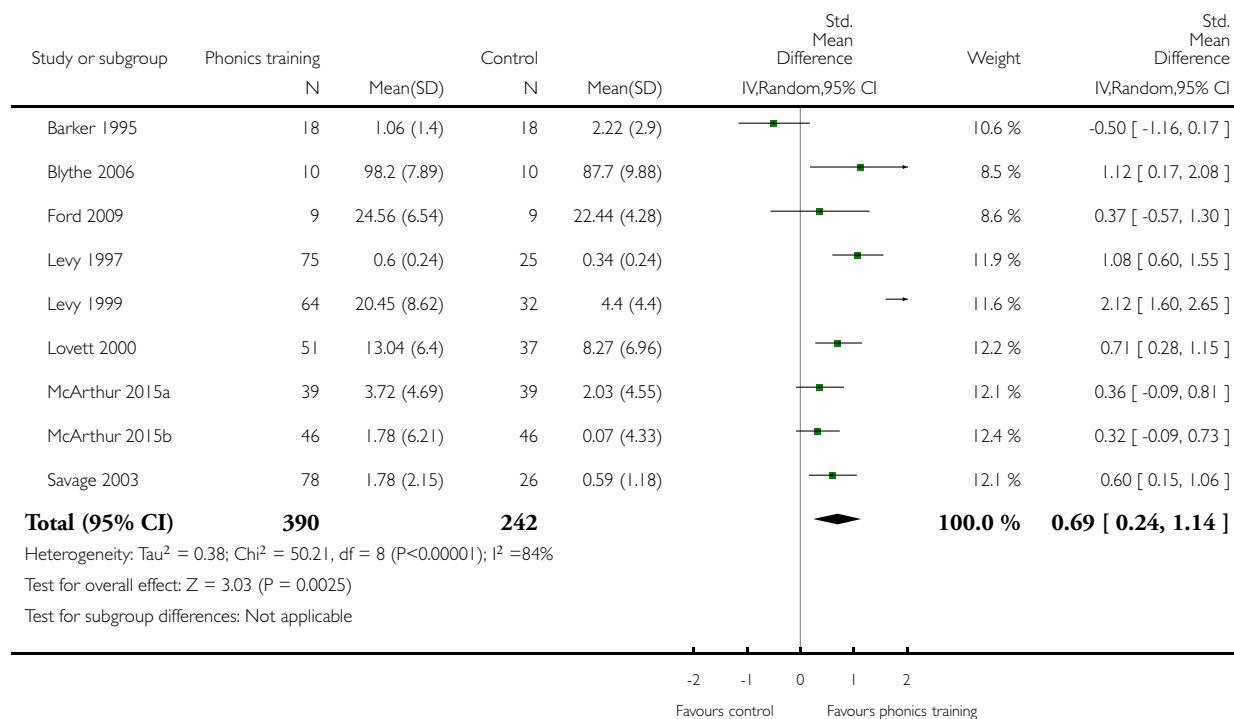


Analysis 4.2. Comparison 4 Phonics training versus control: sensitivity analysis with Hurford 1994 removed (random-effects model), Outcome 2 Non-word reading accuracy.

Review: Phonics training for English-speaking poor readers

Comparison: 4 Phonics training versus control: sensitivity analysis with Hurford 1994 removed (random-effects model)

Outcome: 2 Non-word reading accuracy

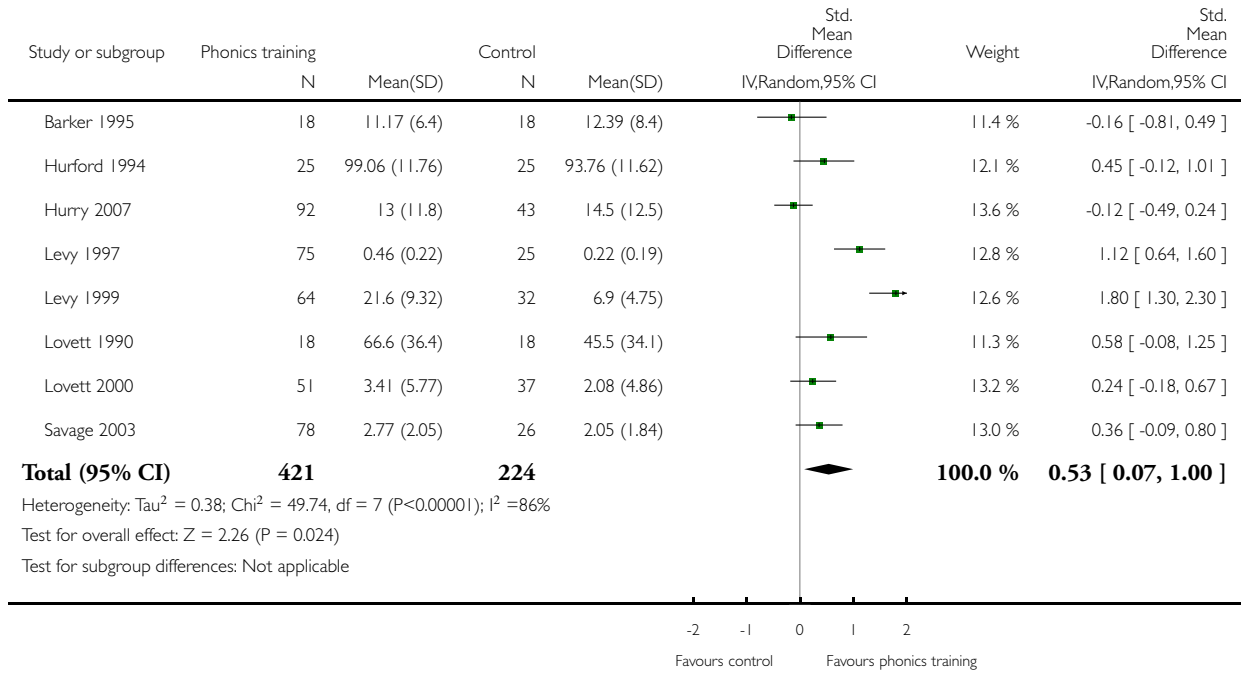


Analysis 5.1. Comparison 5 Phonics training versus control: sensitivity analysis with small studies removed (n < 11), Outcome 1 Mixed/regular word reading accuracy.

Review: Phonics training for English-speaking poor readers

Comparison: 5 Phonics training versus control: sensitivity analysis with small studies removed (n < 11)

Outcome: 1 Mixed/regular word reading accuracy

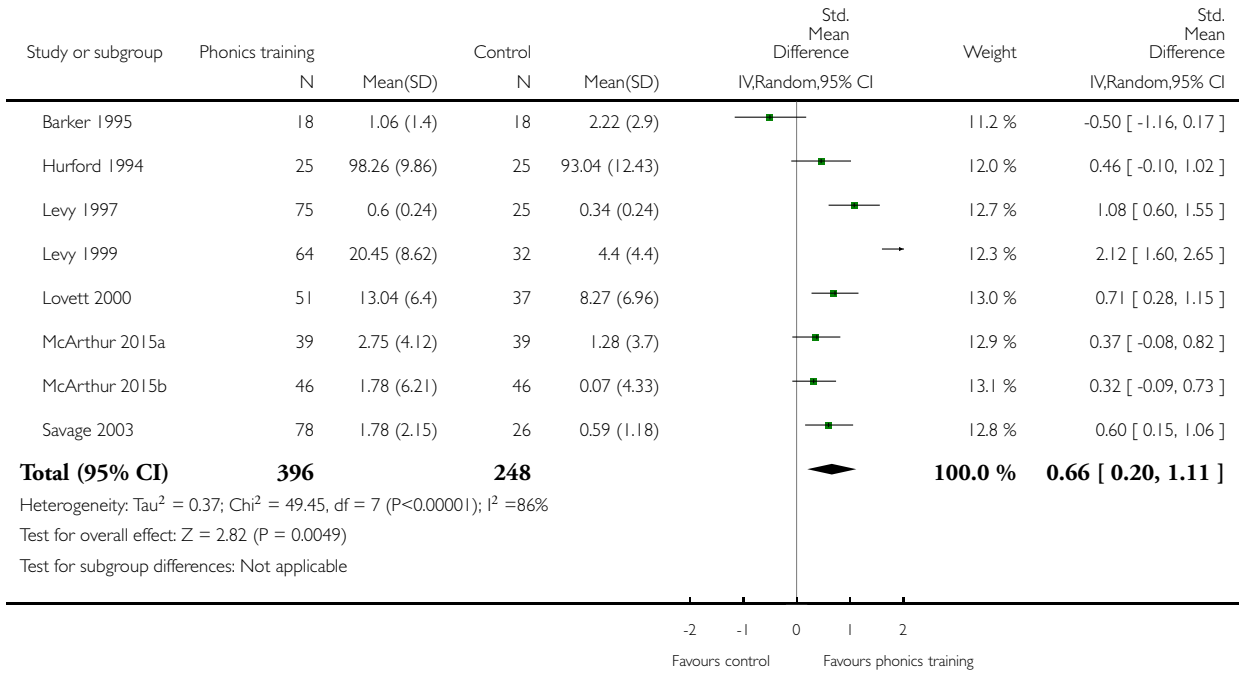


Analysis 5.2. Comparison 5 Phonics training versus control: sensitivity analysis with small studies removed (n < 11), Outcome 2 Non-word reading accuracy.

Review: Phonics training for English-speaking poor readers

Comparison: 5 Phonics training versus control: sensitivity analysis with small studies removed (n < 11)

Outcome: 2 Non-word reading accuracy

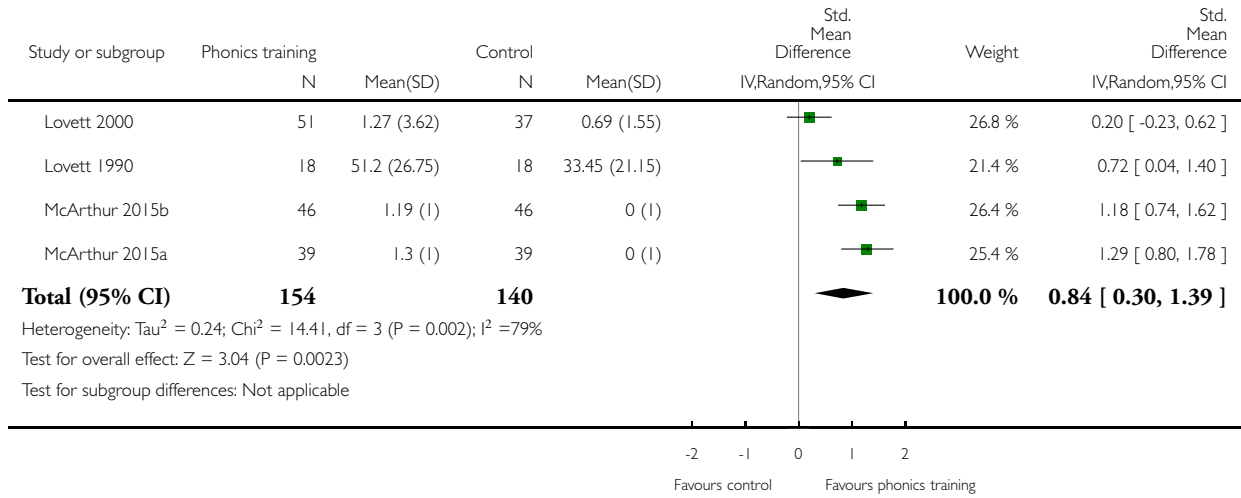


Analysis 5.3. Comparison 5 Phonics training versus control: sensitivity analysis with small studies removed (n < 11), Outcome 3 Irregular word reading accuracy.

Review: Phonics training for English-speaking poor readers

Comparison: 5 Phonics training versus control: sensitivity analysis with small studies removed (n < 11)

Outcome: 3 Irregular word reading accuracy

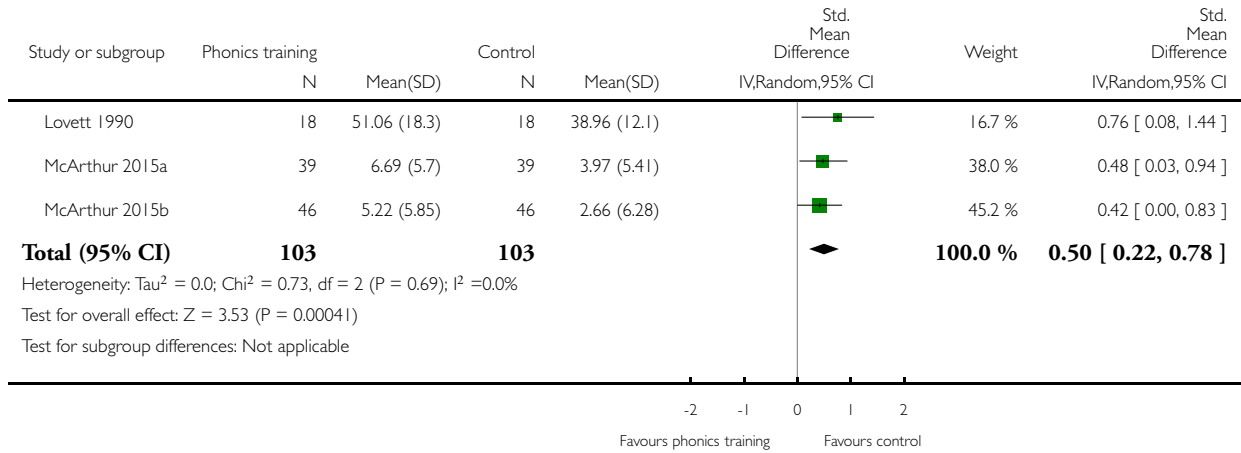


Analysis 5.4. Comparison 5 Phonics training versus control: sensitivity analysis with small studies removed (n < 11), Outcome 4 Mixed/regular word reading fluency.

Review: Phonics training for English-speaking poor readers

Comparison: 5 Phonics training versus control: sensitivity analysis with small studies removed (n < 11)

Outcome: 4 Mixed/regular word reading fluency

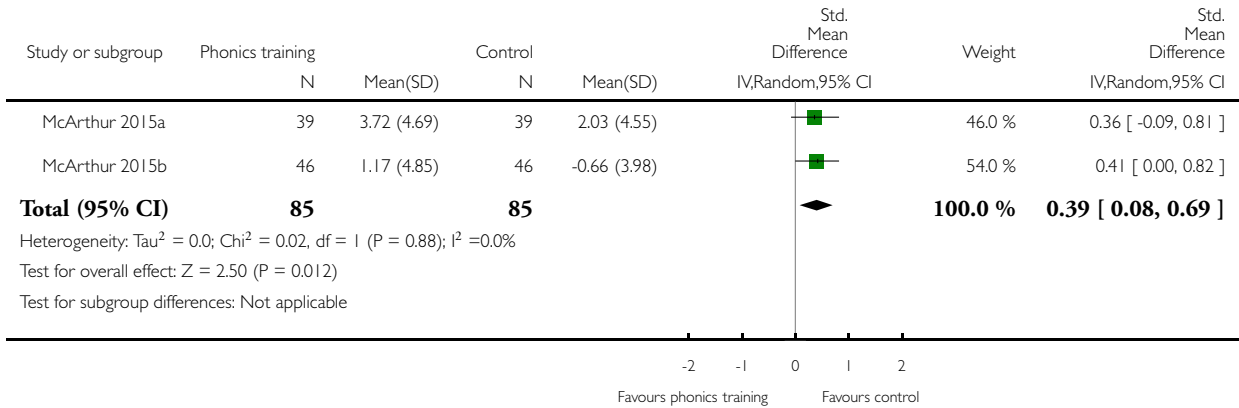


Analysis 5.5. Comparison 5 Phonics training versus control: sensitivity analysis with small studies removed (n < 11), Outcome 5 Non-word reading fluency.

Review: Phonics training for English-speaking poor readers

Comparison: 5 Phonics training versus control: sensitivity analysis with small studies removed (n < 11)

Outcome: 5 Non-word reading fluency

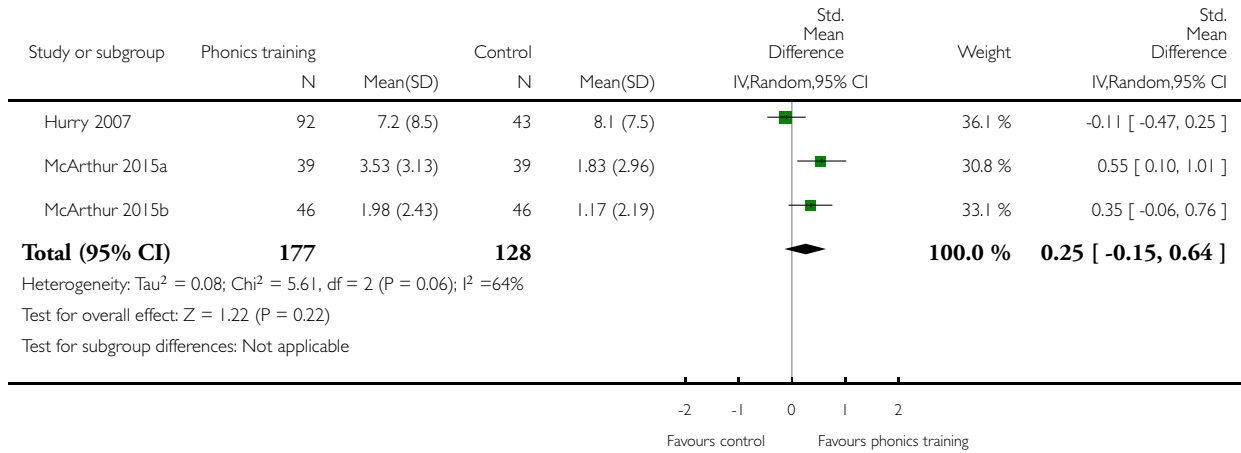


Analysis 5.6. Comparison 5 Phonics training versus control: sensitivity analysis with small studies removed (n < 11), Outcome 6 Reading comprehension.

Review: Phonics training for English-speaking poor readers

Comparison: 5 Phonics training versus control: sensitivity analysis with small studies removed (n < 11)

Outcome: 6 Reading comprehension

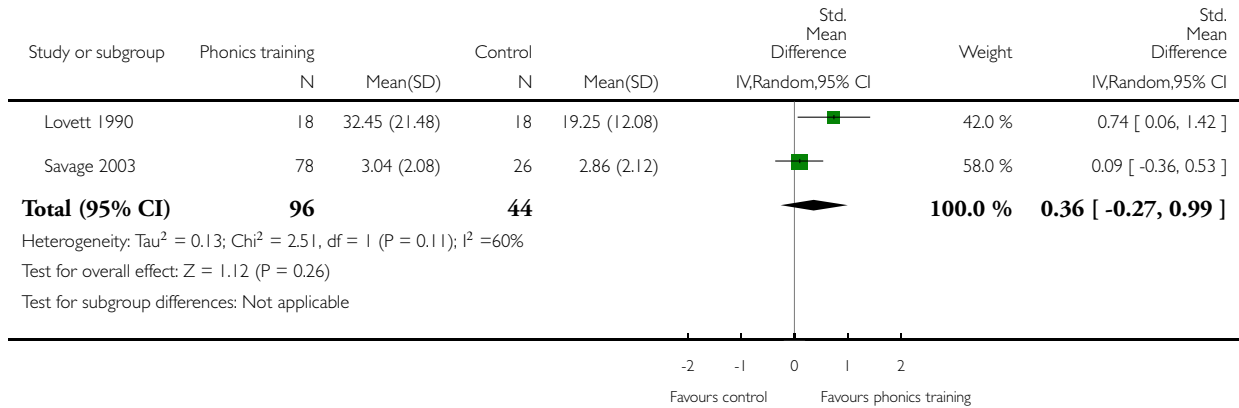


Analysis 5.7. Comparison 5 Phonics training versus control: sensitivity analysis with small studies removed (n < 11), Outcome 7 Spelling.

Review: Phonics training for English-speaking poor readers

Comparison: 5 Phonics training versus control: sensitivity analysis with small studies removed (n < 11)

Outcome: 7 Spelling

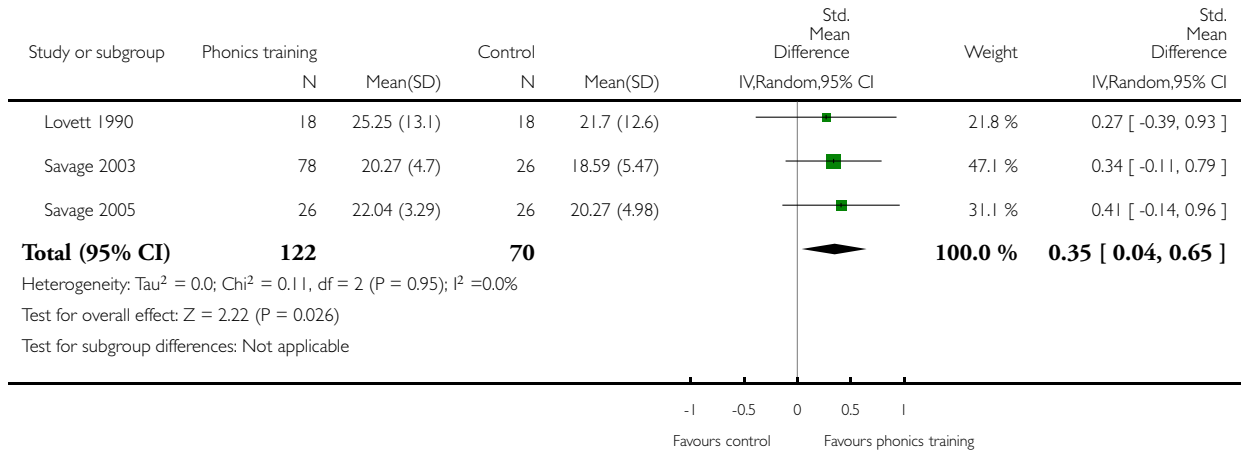


Analysis 5.8. Comparison 5 Phonics training versus control: sensitivity analysis with small studies removed (n < 11), Outcome 8 Letter-sound knowledge.

Review: Phonics training for English-speaking poor readers

Comparison: 5 Phonics training versus control: sensitivity analysis with small studies removed (n < 11)

Outcome: 8 Letter-sound knowledge

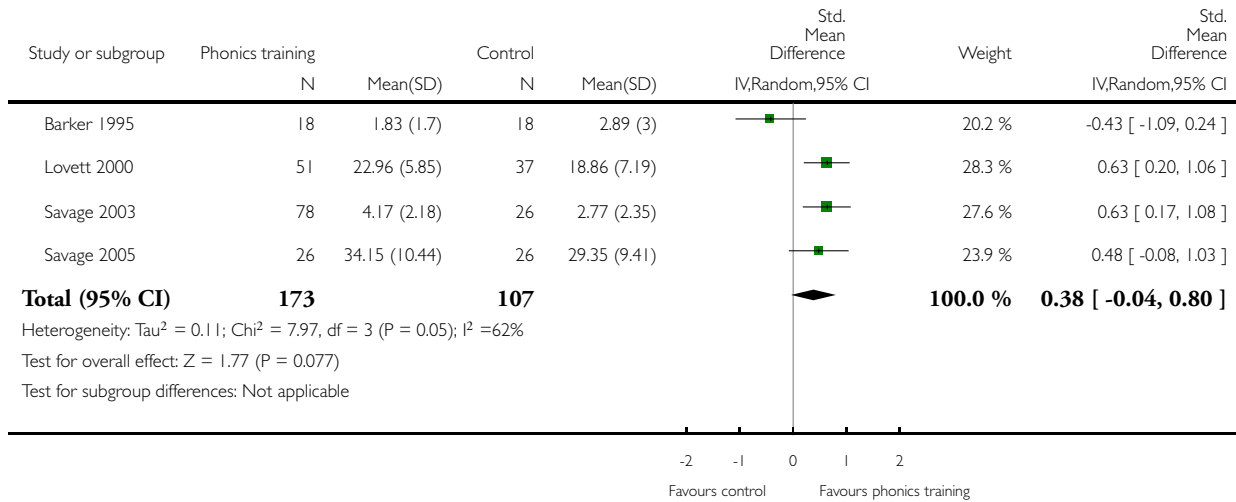


Analysis 5.9. Comparison 5 Phonics training versus control: sensitivity analysis with small studies removed (n < 11), Outcome 9 Phonological output.

Review: Phonics training for English-speaking poor readers

Comparison: 5 Phonics training versus control: sensitivity analysis with small studies removed (n < 11)

Outcome: 9 Phonological output



ADDITIONAL TABLES

Table 1. Tests used by studies to measure outcomes

Outcomes	Tests	References	Studies
Mixed/regular word reading accuracy	Woodcock Johnson Reading Mastery Test Revised: Word Identification	Woodcock 1987	Barker 1995
	Wechsler Individual Achievement Test Second Edition	Wechsler 2001	Blythe 2006
	Woodcock Johnson Psychoeducational Battery Third Edition: Word Identification	Woodcock 2001	Ford 2009
	Woodcock Johnson Reading Mastery Test Revised: Word Identification	Woodcock 1987	Hurford 1994

Table 1. Tests used by studies to measure outcomes (Continued)

	British Ability Scale: Word Reading	Elliot 1983	Hurry 2007
	1 experimental test	Levy 1997	Levy 1997
	1 experimental test	Levy 1999	Levy 1999
	1 experimental test	Lovett 2000	Lovett 2000
	2 experimental tests (trained and untrained - averaged)	Lovett 1990	Lovett 1990
	1 experimental test	Savage 2003	Savage 2003
	Group Reading Assessment and Diagnostic Evaluation Level 1 Version A: Word Recognition Assessment	Williams 2010	Chen 2014
Non-word reading accuracy	Woodcock Johnson Reading Mastery Test Revised: Word Attack	Woodcock 1987	Barker 1995
	Wechsler Individual Achievement Test Second Edition	Wechsler 2001	Blythe 2006
	Woodcock Johnson Psychoeducational Battery Third Edition: Word Attack	Woodcock 2001	Ford 2009
	Woodcock Johnson Reading Mastery Test Revised: Word Attack	Woodcock 1987	Hurford 1994
	1 experimental test	Levy 1997	Levy 1997
	1 experimental test	Levy 1999	Levy 1999
	Woodcock Johnson Reading Mastery Test Revised: Word Attack	Woodcock 1987	Lovett 2000
	1 experimental test	Savage 2003	Savage 2003
	1 experimental test	McArthur 2015a	McArthur 2015a
	1 experimental test	McArthur 2015b	McArthur 2015b

Table 1. Tests used by studies to measure outcomes (Continued)

Irregular word reading accuracy	2 experimental tests (trained and untrained - averaged)	Lovett 1990	Lovett 1990
	1 experimental test	Lovett 2000	Lovett 2000
	1 experimental test	McArthur 2015a	McArthur 2015a
	1 experimental test	McArthur 2015b	McArthur 2015b
Mixed/regular word reading fluency	Test of Word Reading Efficiency: Sight Word subtest	Torgesen 1999b	Ford 2009
	4 experimental tests (regular and irregular - trained and untrained - averaged)	Lovett 1990	Lovett 1990
	Test of Word Reading Efficiency: Sight Word subtest	Torgesen 1999b	McArthur 2015a
	Test of Word Reading Efficiency: Sight Word subtest	Torgesen 1999b	McArthur 2015b
Non-word reading fluency	Test of Word Reading Efficiency: Non-word subtest	Torgesen 1999b	Ford 2009
	Test of Word Reading Efficiency: Non-word subtest	Torgesen 1999b	McArthur 2015a
	Test of Word Reading Efficiency: Non-word subtest	Torgesen 1999b	McArthur 2015b
Reading comprehension	Wechsler Individual Achievement Test Second Edition	Wechsler 2001	Blythe 2006
	Gates-MacGinitie Reading Test Fourth Edition: Comprehension	MacGinitie 2002	Ford 2009
	Neale Analysis of Reading Ability	Neale 1988	Hurry 2007
	Test of Everyday Reading Comprehension	McArthur 2013	McArthur 2015a
	Test of Everyday Reading Comprehension	McArthur 2013	McArthur 2015b

Table 1. Tests used by studies to measure outcomes (Continued)

Spelling	4 experimental tests (regular and irregular - trained and untrained - averaged)	Lovett 1990	Lovett 1990
	1 experimental test	Savage 2003	Savage 2003
	1 experimental test	Chen 2014	Chen 2014
Letter-sound knowledge	2 experimental tests (trained and untrained - aver- aged)	Lovett 1990	Lovett 1990
	1 experimental test	Savage 2003	Savage 2003
	1 experimental test	Savage 2005	Savage 2005
Phonological output (phoneme awareness tasks)	1 experimental test	Barker 1995	Barker 1995
	Goldman Fristoe Woodcock Test of Auditory Discrimina- tion: Sound analysis	Goldman 1974	Lovett 2000
	1 experimental test	Savage 2003	Savage 2003
	1 experimental test	Savage 2005	Savage 2005

Table 2. Effect sizes for random-effects and fixed-effect analyses, and heterogeneity for random-effects analyses

Out- come	N°of studies	N°of parti- pants	Random-effects model			Heterogeneity			Fixed-effect model		
			SMD (95% CI)	Z	P	Chi ²	P	I ² (%)	SMD (95% CI)	Z	P
Mixed/ regu- lar word reading accuracy	11	701	0.51 (0. 13 to 0. 90)	2.59	0.01	52.11	< 0.001	81	0.48 (0.32 to 0.64)	5.78	< 0.001
Non- word reading accuracy	10	682	0.67 (0. 26 to 1. 07)	3.24	0.001	50.72	< 0.001	82	0.68 (0.51 to 0.84)	8.03	< 0.001

Table 2. Effect sizes for random-effects and fixed-effect analyses, and heterogeneity for random-effects analyses (Continued)

Irregular word reading accuracy	4	294	0.84 (0.30 to 1.39)	3.04	0.002	14.41	0.002	79	0.82 (0.58 to 1.07)	6.66	0.002
Mixed/regular word reading fluency	4	224	0.45 (-0.19 to 0.72)	3.33	< 0.001	2.20	0.53	0	0.45 (0.19 to 0.72)	3.33	< 0.001
Non-word reading fluency	3	188	0.39 (0.10 to 0.68)	2.63	0.009	0.02	0.99	0	0.39 (0.10 to 0.68)	2.63	0.009
Reading comprehension	5	343	0.28 (-0.07 to 0.62)	1.54	0.12	8.45	0.08	53	0.23 (0.01 to 0.45)	2.07	0.040
Spelling	3	158	0.47 (-0.07 to 1.01)	1.72	0.09	3.89	0.14	49	0.28 (-0.09 to 0.65)	1.49	0.14
Letter-sound knowledge	3	192	0.35 (0.04 to 0.65)	2.22	0.03	0.11	0.95	0	0.35 (0.04 to 0.65)	2.22	0.03
Phonological output	4	280	0.38 (-0.04 to 0.80)	1.77	0.08	7.97	0.05	62	0.44 (0.19 to 0.70)	3.45	< 0.001

CI: confidence interval; SMD: standardised mean difference.

Table 3. Characteristics of participants in each study

Study	Location	Group N°in analyses	Age	Gender	IQ	Ethnicity	SES	Inclusion criteria	Exclusion criteria	Population
Barker 1995	USA	Intervention: 18 Control: 18	Mean not reported SD not reported Range 6.2-7.8	Not reported	Verbal Mean 16.5 SD 2.36 Range 11-22	Not reported	Not reported	Students nominated by teachers from 2 el-	None stated	First-grade students

Table 3. Characteristics of participants in each study (Continued)

			years					ementary schools who were given a short series of pretests assessing phonological awareness skills and basic word recognition skills. These children were then given further 2 tests and those scoring below the 40th percentile and the 50th percentile on the subsequent test were selected		
Blythe 2006	Australia	Intervention: 10 months Control: 10	Mean 101.5 months SD 17.58 months Range not reported	Male: 75% Female: 25%	FSIQ-2 Mean 100.15 SD 9.38 Range not reported	Not reported	Not reported	Children who received group-based remedial reading instruction at school and were	After referral children completed the WISC-III FSIQ. Those who scored < 20th per-	Dyslexic primary school students

Table 3. Characteristics of participants in each study (Continued)

								re-ferred by a support teacher	centile were ex-cluded	
Chen 2014	Canada	Intervention: 9 Control: 9	Mean 7.06 years SD 0.24 years Range 7-8 years	Male: 39% Female: 61%	Mean 19.79 SD not reported Range not reported	Bilingual speakers of English and French	Not reported	Students considered to be 'at-risk readers' who fall 1 SD below mean on the GRADE (standardised test)	None stated	Second-grade students
Ford 2009	USA	Intervention: 9 Control: 9	Mean 16.18 years SD not reported Range not reported	Male: 55% Female: 45%	Not reported	22% African-American, 67% Hispanic, 11% White	Lower	Students who were enrolled in the remedial reading programme were invited to participate. Below mean reading skills were based on the ISAT	None stated	Teenagers enrolled at an alternative high school, that is, a high school for non-special education students or students at risk of dropping out
Hurford 1994	USA	Intervention: 25 Control: 25	Mean 80.35 months SD not reported Range not reported	Male: 48% Female: 52%	Mean 90.37 SD not reported Range not reported	92.8% white, 6% African-American, 5% Hispanic, 7% Asian-American	Middle	Classification data from Hurford 1993 was used with more relaxed criteria for eligibil-	None stated	Children at risk of reading disability

Table 3. Characteristics of participants in each study (Continued)

								ity, that is standard scores in reading of < 91 were included rather than < 86		
Hurry 2007	UK	Intervention: 92 Control: 43	Mean not reported SD not reported Range 6-6.6 years	Male: 61% Female: 39%	Mean not reported SD not reported Range 92-96	16% spoke English as a second language	42% of the sample were eligible for free school meals.	In 63 schools, the 6 poorest year 2 readers were selected on the basis of their Diagnostic Survey (Clay 1985) performance. Of the 22 schools using Reading Recovery, the poorest scorers were offered intervention	The remaining children, that is, those less poor at reading than those that were selected for the experimental condition, were assigned to a within school condition	Children with reading difficulties
Levy 1997	Canada	Intervention: 75 Control: 25	Mean not reported SD not reported Range 5.9-7.2 years	Male: 48% Female: 52%	Not reported	Not reported	Not reported	Children were given word reading tests, children that read < 7 words on any of the	None stated	All children from Grade 1 and senior kindergarten from 2 schools,

Table 3. Characteristics of participants in each study (Continued)

								screening tests were selected		whose parents consented to their participation
Levy 1999	Canada	Intervention: 64 Control: 32	Mean 7.7 years SD not reported Range not reported	Male: 56% Female: 44%	Non-verbal Experimental group: Mean 10.88 SD not reported Range not reported Control group: Mean 10.65 SD not reported Range not reported	Mixed racial distribution	Covers all SES	Children were given a word identification test (WRAT-3), if they scored < 90 they were given another word identification test (WRMT) and if they read below half a grade below their grade level and read no more than 15 of the training words then they were included in the sample	None stated	17 schools participated in the screening process with permission for participation obtained from the board, schools and a parent or guardian
Lovett 1990	Canada	Intervention: 18	Mean 8.4 years	Male: 70.4%	Verbal Mean 98.	Not reported	Middle	Children had	Children with	Children referred

Table 3. Characteristics of participants in each study (Continued)

		Control: 18	SD 1.6 years Range 7-13 years	Female: 29.6%	4 SD 10.6 Range not reported Performance Mean 106.2 SD 12.6 Range not reported			to score < 25th percentile on at least 4 of 5 reading measures used in the screening test and have at least low mean intelligence	English as a second language, history of extreme hyperactivity, hearing impairment, brain damage, a chronic medical condition, serious emotional disturbance, or attention deficits	to the Learning Disabilities Reading Program
Lovett 2000	Canada	Intervention: 51 Control: 37	Mean 9.9 years SD 1.6 years Range 7-13 years	Male: 68.1% Female: 31.9%	Verbal Mean 92 SD 13.7 Range 58-133 Performance Mean 98.7 SD 14.3 Range 63-136	Not reported	Not reported	Children needed to demonstrate a 'substantial underachievement' on 4 of the 5 reading based screening assessments	None stated	Children with severe reading disabilities that were referred to the Clinical Research Unit for remediation
McArthur 2015a	Australia	Intervention: 39 Control: 39	Mean 9.42 years SD 1.71 years Range 7-12 years	Male: 63.8% Female: 36.2%	Non-verbal Group 1: Mean 97.50 SD 14.16 Range not reported	Not reported	Not reported	Children who scored below the mean range for their age on the Castles	History of neurological or sensory impairment; non-English speakers	Children with reading difficulties

Table 3. Characteristics of participants in each study (Continued)

					Group 2: Mean 95.56 SD 17.12 Range not reported			and Coltheart irregular word reading test and/or non-word reading test		
McArthur 2015b	Australia	Intervention: 46 Control: 46	Group 1: Mean 9.53 years SD 1.51 years Range 7-12 years Group 2: Mean 9.58 years SD 1.45 years Range 7-12 years	Male: 46.3% Female: 53.7%	Non-verbal Group 1: Mean 97.02 SD 15.75 Range not reported Group 2: Mean 95.57 SD 1.65 Range not reported	Not reported	Not reported	Children who scored below the mean range for their age on the Castles and Coltheart irregular word reading test and/or non-word reading test	History of neurological or sensory impairment; non-English speakers	Children with reading difficulties
Savage 2003	UK	Intervention: 78 Control: 26	Mean 5.9 years SD not reported Range 5-6.3 years	Male: 60% Female: 40%	Not reported	Not reported	Not reported	Over 2 sessions a series of reading- and spelling-based assessments were used to find the poorest readers in year 1 of the school. The lowest	A teacher identifying a child as being too immature to deal with working in small groups	Children with the lowest reading performance for their age within a Local Education Authority or School District

Table 3. Characteristics of participants in each study (Continued)

									performers were recruited		
Savage 2005	UK	Intervention: 26 Control: 26	Not reported	Male: 50% Female: 50%	Not reported	Not reported	Lower	Over 2 sessions a series of reading- and spelling-based assessments were used to find the poorest readers in year 1 of the school. The lowest performers were recruited	None stated	Children with the lowest reading performance for their age within a Local Education Authority or School District	

FSIQ: Full Scale Intelligence Quotient; IQ: intelligence quotient; ISAT: Illinois State Achievement Test; SD: standard deviation; SES: socioeconomic status; WISC: Wechsler Intelligence Scale for Children; WRAT: Wide Range Achievement Test; WRMT: Woodcock Reading Mastery Test.

Table 4. Allocation of studies to different subgroups (categories)

Subgroups		Barker 1995	Blythe 2006	Chen 2014	Ford 2009	Hurford 1994	Hurry 2007	Levy 1997	Levy 1999	Lovett 1990	Lovett 2000	McArtl 2015a	McArtl 2015b	Savage 2003	Savage 2005
Training type	Phonics only	X	-	-	-	-	-	X	X	-	-	X	X	-	-
	Phonics + phonem awareness	-	X	-	X	X	X	-	-	-	X	-	-	X	X

Table 4. Allocation of studies to different subgroups (categories) (Continued)

	Phonics + sight words	-	-	X	-	-	-	-	-	X	-	-	-	-	-
Training intensity	< 2 hours/week	X	X	X	X	X	X	X	X	-	-	-	-	X	X
	≥ 2 hours/week	-	-	-	-	-	-	-	-	X	X	X	X	-	-
Training duration	< 3 months	X	X	X	X	-	-	X	X	X	X	X	X	X	X
	≥ 3 months	-	-	-	-	X	X	-	-	-	-	-	-	-	-
Training group size	1	-	X	-	X	X	X	X	X	-	-	X	X	-	-
	≤ 5	X	-	X	-	-	-	-	-	X	X	-	-	X	X
Training administrator	Human	-	-	X	-	-	X	X	X	X	X	-	-	X	X
	Computer	X	X	-	X	X	-	-	-	-	-	X	X	-	-

Table 5. Results of subgroup analyses

Outcome	Subgroups	N° studies	N° participants	Mean effect size			Heterogeneity			Subgroup analyses				
				SMD (95% CI)	Z	P	Chi ²	P	I ² (%)	Chi ²	DF	P	I ² (%)	
Mixed/reg-	Training type	Phonics only	3	232	0.94 (-0.00)	1.79	0.07	21.83	< 0.001	91	-	-	-	-

Table 5. Results of subgroup analyses (Continued)

ular word read- ing accu- racy				09 to 1.97)									
	Phon- ics + phonem aware- ness	6	415	0.17(- 0.04 to 0.37)	1.61	0.11	4.35	0.50	0	-	-	-	-
	Phon- ics + sight word	2	54	0.73 (0. 18 to 1.29)	2.58	0.01	0.61	0.43	0	5.22	2	0.07	61.70
Train- ing in- ten- sity	< 2 hours/ week	9	577	0.54 (0. 06 to 1.02)	2.19	0.03	50.68	< 0. 001	84	-	-	-	-
	≥ 2 hours/ week	2	124	0.34(- 0.02 to 0.70)	1.87	0.06	0.71	0.40	0	0.42	1	0.52	0
Train- ing dura- tion	< 3 months	9	516	0.61 (0. 17 to 1.05)	2.70	0.007	38.25	< 0. 001	79	-	-	-	-
	≥ 3 months	2	185	0.12(- 0.43 to 0.67)	0.42	0.67	2.80	0.09	64	1.84	1	0.17	45.80
Train- ing group size	1	6	419	0.62(- 0.06 to 1.29)	1.78	0.07	44.35	< 0. 001	89	-	-	-	-
	≤ 5	5	282	0.33 (0. 04 to 0.61)	2.24	0.02	4.94	0.29	19	0.59	1	0.44	0
Train- ing ad-	Hu- man	7	577	0.70 (0.	2.57	0.01	46.63	< 0. 001	87	-	-	-	-

Table 5. Results of subgroup analyses (Continued)

	minis- trator				17 to 1.23)									
		Com- puter	4	124	0.18 (- 0. 20 to 0.51)	1.00	0.32	2.01	0.57	0	2.51	1	0.11	60.20
Non- word read- ing ac- curacy	Train- ing type	Phon- ics only	5	402	0.69 (- 0. 08 to 1.66)	1.75	0.08	48.66	< 0. 001	92	-	-	-	-
		Phon- ics + phonem aware- ness	5	280	0. 63 (0. 38 to 0.88)	4.86	< 0. 001	1.84	0.77	0	0.02	1	0.89	0
	Train- ing group size	1	7	454	0. 83 (0. 31 to 1.36)	3.10	0.002	37.34	< 0. 001	84	-	-	-	-
		≤ 5	3	228	0.32 (- 0. 32 to 0.96)	0.97	0.33	9.64	0.008	79	1.47	1	0.23	31.80
	Train- ing ad- minis- trator	Hu- man	4	388	1. 12 (0. 48 to 1.76)	3.42	< 0. 001	22.23	< 0. 001	87	-	-	-	-
		Com- puter	6	294	0.31 (- 0. 02 to 0.64)	1.85	0.06	8.81	0.12	43	4.84	1	0.03	79.40

CI: confidence interval; DF: degrees of freedom; SMD: standardised mean difference.

Table 6. Quality of evidence ratings for primary and secondary outcomes (based on Ryan 2016)

Outcome	Study quality RCT = high Non-RCT = low	Risk of bias ^a No = 0 Serious = -1 Very serious = -2	Inconsistency ^b No = 0 Serious = -1 Very serious = -2	Indirectness No = 0 Serious = -1 Very serious = -2	Imprecision ^e No = 0 Serious = -1 Very serious = -2	Publication bias ^f Undetected = 0 Strongly suspected = -1	Other Large effect = + 1 Dose effect = + 1 No plausible confound = + 1	GRADE
Mixed/regular word reading accuracy	High	No = 0	No = 0 ^c	No = 0	Very serious = -2	Undetected = 0	-	Low
Non-word reading accuracy	High	No = 0	No = 0 ^c	No = 0	Very serious = -2	Undetected = 0	-	Low
Irregular word reading accuracy	High	No = 0	No = 0 ^c	No = 0	Very serious = -2	Undetected = 0	Large effect = + 1	Moderate
Mixed/regular word reading fluency	High	No = 0	No = 0 ^d	No = 0	Serious = -1	Undetected = 0	-	Moderate
Non-word reading fluency	High	No = 0	No = 0 ^d	No = 0	Serious = -1	Undetected = 0	-	Moderate
Reading comprehension	High	No = 0	No = 0 ^d	No = 0	Very serious = -2	Undetected = 0	-	Low
Spelling	High	No = 0	No = 0 ^d	No = 0	Very serious = -2	Undetected = 0	-	Low
Letter-sound knowledge	High	No = 0	No = 0 ^d	No = 0	Very serious = -2	Undetected = 0	-	Low
Phonological output	High	No = 0	No = 0 ^d	No = 0	Very serious = -2	Undetected = 0	-	Low

^aJudged 'no' if 75% + studies contributing to an outcome are low in majority of biases. Judged 'serious' if 50% to 74% of studies contributing to an outcome are low in majority of biases. Judged 'very serious' if fewer than 50% studies contributing to an outcome are low in majority of biases. See 'Risk of bias' [Figure 1](#) and 'Risk of bias' tables for bias ratings for each study.

^bJudged 'no' if I^2 less than 70%^d OR I^2 greater than 70% *but* assessment of heterogeneity analysis suggests it did not affect the reliability of results^c (see [Subgroup analysis and investigation of heterogeneity](#)). Judged 'serious' if $I^2 = 70\%$ to 85%; judged 'very serious' if I^2 greater than 85%.

^eJudged 'no' if confidence interval 0 to 0.3. Judged 'serious' if confidence interval 0.3 to 0.6. Judged 'very serious' if confidence interval 0.6 + (Schünemann 2011b).

^fJudged 'undetected' if funnel plot done on more than 10 studies (Sterne 2011), and no bias detected. Judged 'unsuspected' if funnel plot not constructed (too few studies) but bias not strongly suspected. Judged 'strong suspected' if funnel plot not possible (too few studies) and bias strongly suspected.

APPENDICES

Appendix I. Search strategies

Cochrane Central Register of Controlled Trials (CENTRAL), in the Cochrane Library

- #1 MeSH descriptor Reading, this term only
- #2 MeSH descriptor Dyslexia, this term only
- #3 (read* near/3 disorder*)
- #4 (read* near/3 (abilit* or disab*))
- #5 (read* near/3 impair*)
- #6 (read* near/3 defic*)
- #7 (read* near/3 delay*)
- #8 (read* near/3 dysfunction*)
- #9 (read* near/3 comprehen*)
- #10 (read* near/3 accuracy)
- #11 (poor* near/3 read*)
- #12 ((dysfluent or dysfluenc* or fluent or fluenc*) near/3 read*)
- #13 (slow* near/3 read*)
- #14 (remedial near/3 read*)
- #15 dyslex*
- #16 (word NEXT blind* or wordblind*)
- #17 (#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16)
- #18 MeSH descriptor Phonetics, this term only
- #19 phonics
- #20 phonem*
- #21 phonolog*
- #22 graphem*
- #23 (lettersound* or letter NEXT sound*)
- #24 letter NEXT identif*
- #25 (sight NEXT word*)
- #26 MeSH descriptor Remedial Teaching, this term only
- #27 (remedial near/3 (teach* or method* or program*))
- #28 (#18 OR #19 OR #20 OR #21 OR #22 OR #23 OR #24 OR #25 OR #26 OR #27)
- #29 (#17 AND #28)

MEDLINE Ovid

- 1 Reading/
- 2 (read\$ adj3 disorder\$).tw.
- 3 (read\$ adj3 (abilit\$ or disab\$)).tw.
- 4 (read\$ adj3 impair\$).tw.
- 5 (read\$ adj3 defic\$).tw.
- 6 (read\$ adj3 delay\$).tw.
- 7 (read\$ adj3 dysfunction\$).tw.
- 8 (read\$ adj3 comprehen\$).tw.
- 9 (read\$ adj3 accuracy).tw.
- 10 (poor\$ adj3 read\$).tw.
- 11 ((dysfluent or dysfluenc\$ or fluent or fluenc\$) adj3 read\$).tw.
- 12 (slow\$ adj3 read\$).tw.
- 13 (remedial adj3 read\$).tw.
- 14 dyslexia/
- 15 dyslex\$.tw.
- 16 (word-blind\$ or wordblind\$).tw.
- 17 or/1-16
- 18 phonics.tw.
- 19 phonem\$.tw.
- 20 phonolog\$.tw.
- 21 graphem\$.tw.
- 22 (lettersound\$ or letter-sound\$).tw.
- 23 letter identif\$.tw.
- 24 (sight word\$ or sight-word\$).tw.
- 25 Phonetics/
- 26 Remedial Teaching/
- 27 (remedial adj3 (teach\$ or method\$ or program\$)).tw.
- 28 or/18-27
- 29 17 and 28
- 30 randomized controlled trial.pt.
- 31 controlled clinical trial.pt.
- 32 randomi#ed.ab.
- 33 placebo\$.ab.
- 34 drug therapy.fs.
- 35 randomly.ab.
- 36 trial.ab.
- 37 groups.ab.
- 38 or/30-37
- 39 exp animals/ not humans.sh.
- 40 38 not 39
- 41 29 and 40

MEDLINE In-process & Other Non-Indexed Citations Ovid

- 1 dyslex\$.tw,kf.
- 2 (read\$ adj3 disorder\$).tw,kf.
- 3 (read\$ adj3 (abilit\$ or disab\$)).tw,kf.
- 4 (read\$ adj3 impair\$).tw,kf.
- 5 (read\$ adj3 defic\$).tw,kf.
- 6 (read\$ adj3 delay\$).tw,kf.
- 7 (read\$ adj3 dysfunction\$).tw,kf.
- 8 (poor\$ adj3 read\$).tw,kf.

9 (dysfluen\$ adj3 read\$).tw,kf.
 10 (slow\$ adj3 read\$).tw,kf.
 11 (word-blind\$ or wordblind\$).tw,kf.
 12 or/1-11
 13 phonetic\$.tw,kf.
 14 phonic\$.tw,kf.
 15 phonem\$.tw,kf.
 16 phonolog\$.tw,kf.
 17 graphem\$.tw,kf.
 18 (lettersound\$ or letter-sound\$).tw,kf.
 19 letter identif\$.tw,kf.
 20 (sight word\$ or sight-word\$).tw,kf.
 21 (remedial adj3 (teach\$ or train\$ or method\$ or program\$)).tw,kf.
 22 or/13-21
 23 12 and 22
 24 (random\$ or trial\$ or control\$ or group\$ or placebo\$ or blind\$ or prospectiv\$ or longitudinal\$ or meta-analys\$ or systematic review\$).tw,kf.

MEDLINE Epub Ahead of Print Ovid

1 dyslex\$.tw,kf.
 2 (read\$ adj3 disorder\$).tw,kf.
 3 (read\$ adj3 (abilit\$ or disab\$)).tw,kf.
 4 (read\$ adj3 impair\$).tw,kf.
 5 (read\$ adj3 defici\$).tw,kf.
 6 (read\$ adj3 delay\$).tw,kf.
 7 (read\$ adj3 dysfunction\$).tw,kf.
 8 (poor\$ adj3 read\$).tw,kf.
 9 (dysfluen\$ adj3 read\$).tw,kf.
 10 (slow\$ adj3 read\$).tw,kf.
 11 (word-blind\$ or wordblind\$).tw,kf.
 12 or/1-11
 13 phonetic\$.tw,kf.
 14 phonic\$.tw,kf.
 15 phonem\$.tw,kf.
 16 phonolog\$.tw,kf.
 17 graphem\$.tw,kf.
 18 (lettersound\$ or letter-sound\$).tw,kf.
 19 letter identif\$.tw,kf.
 20 (sight word\$ or sight-word\$).tw,kf.
 21 (remedial adj3 (teach\$ or train\$ or method\$ or program\$)).tw,kf.
 22 or/13-21
 23 12 and 22
 24 (random\$ or trial\$ or control\$ or group\$ or placebo\$ or blind\$ or prospectiv\$ or longitudinal\$ or meta-analys\$ or systematic review\$).tw,kf.

Embase Ovid

1 reading/
 2 dyslexia/
 3 (read\$ adj3 disorder\$).tw.
 4 (read\$ adj3 (abilit\$ or disab\$)).tw.
 5 (read\$ adj3 impair\$).tw.

6 (read\$ adj3 defic\$).tw.
 7 (read\$ adj3 delay\$).tw.
 8 (read\$ adj3 dysfunction\$).tw.
 9 (read\$ adj3 comprehen\$).tw.
 10 (read\$ adj3 accuracy).tw.
 11 (poor\$ adj3 read\$).tw.
 12 (read\$ adj3 (fluent or fluenc\$ or dysfluent or dysfluenc\$)).tw.
 13 (slow\$ adj3 read\$).tw.
 14 (remedial adj3 read\$).tw.
 15 dyslex\$.tw.
 16 (word-blind\$ or wordblind\$).tw.
 17 or/1-16
 18 phonics.tw.
 19 phonem\$.tw.
 20 phonolog\$.tw.
 21 graphem\$.tw.
 22 (lettersound\$ or letter-sound\$).tw.
 23 letter identif\$.tw.
 24 (sight word\$ or sight-word\$).tw.
 25 Phonetics/
 26 (remedial adj3 (teach\$ or method\$ or program\$)).tw.
 27 or/18-26
 28 17 and 27
 29 exp Clinical trial/
 30 Randomized controlled trial/
 31 Randomization/
 32 Single blind procedure/
 33 Double blind procedure/
 34 Crossover procedure/
 35 Placebo/
 36 Randomi#ed.tw.
 37 RCT.tw.
 38 (random\$ adj3 (allocat\$ or assign\$)).tw.
 39 randomly.ab.
 40 groups.ab.
 41 trial.ab.
 42 ((singl\$ or doubl\$ or trebl\$ or tripl\$) adj3 (blind\$ or mask\$)).tw.
 43 Placebo\$.tw.
 44 Prospective study/
 45 (crossover or cross-over).tw.
 46 prospective.tw.
 47 or/29-46
 48 28 and 47

Cochrane Database of Systematic Reviews (CDSR), part of the Cochrane Library

#1 MeSH descriptor Reading, this term only
 #2 MeSH descriptor Dyslexia, this term only
 #3 (read* near/3 disorder*)
 #4 (read* near/3 (abilit* or disab*))
 #5 (read* near/3 impair*)
 #6 (read* near/3 defic*)
 #7 (read* near/3 delay*)

#8 (read* near/3 dysfunction*)
 #9 (read* near/3 comprehen*)
 #10 (read* near/3 accuracy)
 #11 (poor* near/3 read*)
 #12((dysfluent or dysfluenc* or fluent or fluenc*) near/3 read*)
 #13(slow* near/3 read*)
 #14(remedial near/3 read*)
 #15 dyslex*
 #16(word NEXT blind* or wordblind*)
 #17(#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16)
 #18 MeSH descriptor Phonetics, this term only
 #19 phonics
 #20 phonem*
 #21 phonolog*
 #22 graphem*
 #23 (lettersound* or letter NEXT sound*)
 #24 letter NEXT identif*
 #25 (sight NEXT word*)
 #26 MeSH descriptor Remedial Teaching, this term only
 #27 (remedial near/3 (teach* or method* or program*))
 #28 (#18 OR #19 OR #20 OR #21 OR #22 OR #23 OR #24 OR #25 OR #26 OR #27)
 #29 (#17 AND #28)

Database of Abstracts of Reviews of Effects (DARE), part of the Cochrane Library

#1 MeSH descriptor Reading, this term only
 #2 MeSH descriptor Dyslexia, this term only
 #3 (read* near/3 disorder*)
 #4 (read* near/3 (abilit* or disab*))
 #5 (read* near/3 impair*)
 #6 (read* near/3 defic*)
 #7 (read* near/3 delay*)
 #8 (read* near/3 dysfunction*)
 #9 (read* near/3 comprehen*)
 #10 (read* near/3 accuracy)
 #11 (poor* near/3 read*)
 #12((dysfluent or dysfluenc* or fluent or fluenc*) near/3 read*)
 #13(slow* near/3 read*)
 #14(remedial near/3 read*)
 #15 dyslex*
 #16(word NEXT blind* or wordblind*)
 #17(#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16)
 #18 MeSH descriptor Phonetics, this term only
 #19 phonics
 #20 phonem*
 #21 phonolog*
 #22 graphem*
 #23 (lettersound* or letter NEXT sound*)
 #24 letter NEXT identif*
 #25 (sight NEXT word*)
 #26 MeSH descriptor Remedial Teaching, this term only
 #27 (remedial near/3 (teach* or method* or program*))
 #28 (#18 OR #19 OR #20 OR #21 OR #22 OR #23 OR #24 OR #25 OR #26 OR #27)

#29 (#17 AND #28)

ERIC (Education Resources Information Center)

ERIC EBSCOhost searched 2012 onwards

S1 DE "Reading" OR DE "Basal Reading" OR DE "Beginning Reading" OR DE "Content Area Reading" OR DE "Corrective Reading" OR DE "Critical Reading" OR DE "Directed Reading Activity" OR DE "Early Reading" OR DE "Functional Reading" OR DE "Independent Reading" OR DE "Individualized Reading" OR DE "Music Reading" OR DE "Oral Reading" OR DE "Reading Aloud to Others" OR DE "Recreational Reading" OR DE "Silent Reading" OR DE "Speed Reading" OR DE "Story Reading" OR DE "Sustained Silent Reading"

S2 DE "Dyslexia"

S3 DE "Reading Difficulties" OR DE "Reading Fluency" OR DE "Reading Improvement"

S4 (read* N3 (abilit* or accuracy or comprehen* or defic* or delay* or disab* or disorder* or dysfunction*))

S5 ((dysfluent or dysfluenc* or fluent or fluenc*) N3 read*)

S6 (slow* OR poor* OR remedial*) N5 read*)S7wordblind* OR "word blind*" OR word-blind*

S8 dyslex*

S9 S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8

S10 DE Phonics

S11 DE "Phonological Awareness"

S12 DE "Phonemic Awareness"

S13 phonic* OR phonem* OR phonolog* OR grapheme*

S14 "letter identif*"

S15 DE "Remedial Programs" OR DE "Remedial Reading"

S16 (remedial N3 (teach* or train* or method* or program*))

S17 S10 OR S11 OR S12 OR S13 OR S14 OR S15 OR S16

S18 S9 AND S17

S19 DE "Meta Analysis" OR DE "Evaluation Research" OR DE "Control Groups" OR DE "Experimental Groups" OR DE "Longitudinal Studies" OR DE "Followup Studies" OR DE "Program Effectiveness" OR DE "Program Evaluation"

S20 TI (random* or trial* or experiment* or PROSPECTIVE* OR longitudinal or BLIND* or CONTROL*) OR AB (random* or trial* or experiment* or PROSPECTIVE* OR longitudinal or BLIND* or CONTROL*)

S21 S19 OR S20

S22 S18 AND S21

ERIC Proquest searched up to 2012

Searched for:((SU.EXACT(EXPLODE("Basal Reading" OR "Beginning Reading" OR "Content Area Reading" OR "Corrective Reading" OR "Critical Reading" OR "Directed Reading Activity" OR "Early Reading" OR "Functional Reading" OR "Independent Reading" OR "Individualized Reading" OR "Music Reading" OR "Oral Reading" OR "Reading" OR "Reading Aloud to Others" OR "Reading Fluency" OR "Recreational Reading" OR "Remedial Reading" OR "Silent Reading" OR "Speed Reading" OR "Story Reading" OR "Sustained Silent Reading") OR SU.EXACT("Dyslexia") OR SU.EXACT("Reading Difficulties") OR ((slow* OR poor* OR remedial*) NEAR/5 read*) OR wordblind* OR "word blind*" OR word-blind*) AND (SU.EXACT("Phonics") OR SU.EXACT("Phonological Awareness") OR SU.EXACT("Phonemic Awareness") OR phonic* OR phonem* OR phonolog* or grapheme* OR "letter identif*"))

ERIC DialogDatastar searched up to May 2011

"(((READING#.W..DE.) OR ((READ\$3 NEAR (DISORDERS\$ OR ABILITY OR DISABILIT\$3 OR IMPAIR\$4 OR DEFIC\$5 OR DELAY\$2 OR DYSFUNCTION\$1)).TI,AB.) OR (((DYSFLUEN\$ OR FLUEN\$) NEAR READ\$3) .TI,AB.) OR (((SLOW\$ OR POOR\$ OR REMEDIAL\$) NEAR READ\$) .TI,AB.) OR (DYSLEXIA.W..DE.) OR (READING-DIFFICULTIES.DE.) OR ((WORDBLIND\$ OR WORD-BLIND\$ OR WORD ADJ BLIND\$) .TI,AB.)) AND ((PHONICS.W..DE. OR PHONEMIC-AWARENESS.DE. OR PHONOLOGICAL-AWARENESS.DE.) OR ((PHONIC\$ OR PHONEM\$ OR PHONOLOG\$) .TI,AB.) OR (GRAPHEME\$.TI,AB.) OR ((LETTER ADJ IDENTIF\$) .TI,AB.) OR ((SIGHT ADJ WORD\$ OR SIGHT-WORD\$ OR

SIGHTWORD\$) .TI,AB.) OR ((REMEDIAL ADJ READING) .TI,AB.) OR (REMEDIAL-READING.DE.) OR ((REMEDIAL NEAR (TEACH\$ OR METHOD\$ OR PROGRAM\$)) .TI,AB.) OR ((LETTERSOUND\$ OR LETTER-SOUND\$ OR LETTER ADJ SOUND\$) .TI,AB.))

PsycINFO

PsycINFO Ovid searched from 2012 onwards

- 1 reading/ or oral reading/ or remedial reading/ or silent reading/
- 2 reading ability/ or reading achievement/ or reading comprehension/
- 3 reading development/ or reading disabilities/ or dyslexia/ or reading speed/
- 4 sight vocabulary/ or word recognition/
- 5 (read\$ adj3 disorder\$.mp.
- 6 (read\$ adj3 (abilit\$ or disab\$)).mp.
- 7 (read\$ adj3 impair\$.mp.
- 8 (read\$ adj3 defic\$.mp.
- 9 (read\$ adj3 delay\$.mp.
- 10 (read\$ adj3 dysfunction\$.mp.
- 11 (read\$ adj3 comprehen\$.mp.
- 12 (read\$ adj3 accuracy).mp.
- 13 (poor\$ adj3 read\$.mp.
- 14 ((dysfluent or dysfluenc\$ or fluent or fluenc\$) adj3 read\$.mp.
- 15 (slow\$ adj3 read\$.mp.
- 16 (remedial adj3 read\$.mp.
- 17 dyslex\$.mp.
- 18 (word-blind\$ or wordblind\$.mp.
- 19 or/1-18
- 20 phonemes/ or phonetics/ or phonics/ or phonological awareness/ or phonology/
- 21 phonics.mp.
- 22 phonem\$.mp.
- 23 phonolog\$.mp.
- 24 graphem\$.mp.
- 25 (lettersound\$ or letter-sound\$.mp.
- 26 letter identif\$.mp.
- 27 (sight word\$ or sight-word\$.mp.
- 28 Remedial Reading/
- 29 (remedial adj3 (teach\$ or method\$ or program\$)).mp.
- 30 or/20-29
- 31 clinical trials/
- 32 (randomis* or randomiz*).tw.
- 33 (random\$ adj3 (allocat\$ or assign\$)).tw.
- 34 ((clinic\$ or control\$) adj trial\$.tw.
- 35 ((singl\$ or doubl\$ or trebl\$ or tripl\$) adj3 (blind\$ or mask\$)).tw.
- 36 (crossover\$ or "cross over\$").tw.
- 37 random sampling/
- 38 Experiment Controls/
- 39 Placebo/
- 40 placebo\$.tw.
- 41 exp program evaluation/
- 42 treatment effectiveness evaluation/
- 43 ((effectiveness or evaluat\$) adj3 (stud\$ or research\$)).tw.
- 44 or/31-43

45 19 and 30 and 44

PsycINFO EBSCOhost searched up to 2011

S45 S30 and S44
S44 S31 or S32 or S33 or S34 or S35 or S36 or S37 or S38 or S39 or S40 or S41 or S42 or S43
S43 (evaluation N3 stud* or evaluation N3 research*)
S42 (effectiveness N3 stud* or effectiveness N3 research*)
S41 DE "Placebo" or DE "Evaluation" or DE "Program Evaluation" OR DE "Educational Program Evaluation" OR DE "Mental Health Program Evaluation" OR DE "Treatment effectiveness evaluation"
S40 (DE "Random Sampling" or DE "Clinical Trials") or (DE "Experiment Controls")
S39 placebo*
S38 crossover* or cross-over* or cross over*
S37 (tripl* N3 mask*) or (tripl* N3 blind*)
S36 (trebl* N3 mask*) or (trebl* N3 blind*)
S35 (doubl* N3 mask*) or (doubl* N3 blind*)
S34 (singl* N3 mask*) or (singl* N3 blind*)
S33 (clinic* N3 trial*) or (control* N3 trial*)
S32 (random* N3 allocat*) or (random* N3 assign*)
S31 randomis* or randomiz*
S30 S18 and S29
S29 S19 or S20 or S21 or S22 or S23 or S24 or S25 or S26 or S27 or S28
S28 (remedial N3 teach*) or (remedial* N3 method*) or (remedial* N3 program*)
S27 DE "Remedial Reading"
S26 sight word* or sight-word* or sightword*
S25 letter identif*
S24 lettersound* or letter-sound* or letter sound*
S23 graphem*
S22 phonolog*
S21 phonem*
S20 phonics
S19 ((DE "Phonics") OR (DE "Phonology")) OR (DE "Phonemes") OR (DE "Phonetics") OR (DE "Phonological Awareness")
S18 S1 or S2 or S3 or S4 or S5 or S6 or S7 or S8 or S9 or S10 or S11 or S12 or S13 or S14 or S15 or S16 or S17
S17 (word-blind* or wordblind* or word blind*)
S16 DE "Dyslexia" or dyslex*
S15 (remedial N3 read*)
S14 (slow* N3 read*) or (poor* N3 read*)
S13 read* N3 comprehen*
S12 read* N3 accurac*
S11 dysfluent N3 read* or dysfluenc* N3 read* or fluent* N3 read* or fluenc* N3 read*
S10 (read* N3 dysfunction*)
S9 (read* N3 delay*)
S8 (read* N3 defic*)
S7 (read* N3 impair*)
S6 read* N3 disab* or read* N3 abilit*
S5 (read* N3 disorder*)
S4 DE "Sight Vocabulary" OR DE "Word Recognition"
S3 DE "Reading Disabilities" OR DE "Dyslexia" OR DE "Reading Speed" OR DE "Reading Development"
S2 DE "Reading Ability" OR DE "Reading Achievement" OR DE "Reading Comprehension"
S1 DE "Reading" OR DE "Oral Reading" OR DE "Remedial Reading" OR DE "Silent Reading"

CINAHL Plus EBSCOhost (Cumulative Index to Nursing and Allied Health Literature)

S50 S31 and S49
 S49 S32 or S33 or S34 or S35 or S36 or S37 or S38 or S39 or S40 or S41 or S42 or S43 or S44 or S45 or S46 or S47 or S48
 S48 (MH "Evaluation Research") OR (MH "Summative Evaluation Research") OR (MH "Program Evaluation")
 S47 (MH "Treatment Outcomes")
 S46 (MH "Comparative Studies")
 S45 (evaluat* study or evaluat* research) or (effectiv* study or effectiv* research) or (prospectiv* study or prospectiv* research) or
 (follow-up study or follow-up research)
 S44 "cross over*"

S43 crossover*
 S42 (MH "Crossover Design") or (MH "Prospective Studies+")
 S41 (tripl* N3 mask*) or (tripl* N3 blind*)
 S40 (trebl* N3 mask*) or (trebl* N3 blind*)
 S39 (doubl* N3 mask*) or (doubl* N3 blind*)
 S38 (singl* N3 mask*) or (singl* N3 blind*)
 S37 (clinic* N3 trial*) or (control* N3 trial*)
 S36 (random* N3 allocat*) or (random* N3 assign*)
 S35 randomis* or randomiz*
 S34 (MH "Meta Analysis")
 S33 (MH "Clinical Trials+")
 S32 MH random assignment
 S31 S18 and S30
 S30 S19 or S20 or S21 or S22 or S23 or S24 or S25 or S26 or S27 or S28 or S29
 S29 remedial N3 teach* or remedial N3 method* or remedial N3 program*
 S28 (MH "Remedial Teaching")
 S27 sight word* or sight-word* or sightword*
 S26 letter identif*
 S25 lettersound* or letter-sound* or letter sound*
 S24 graphem*
 S23 phonolog*
 S22 phonem*
 S21 phonics
 S20 (MH "Phonetics+")
 S19 (MH "Phonology")
 S18 S1 or S2 or S3 or S4 or S5 or S6 or S7 or S8 or S9 or S10 or S11 or S12 or S13 or S14 or S15 or S16 or S17
 S17 word-blind* or wordblind* or word blind*
 S16 dyslex*
 S15 (remedial N3 read*)
 S14 (slow* N3 read*)
 S13 (slow* N3 read*)
 S12 (poor* N3 read*)
 S11 (read* N3 accur*)
 S10 (read* N3 comprehen*)
 S9 (read* N3 fluent) or (read* N3 fluenc*) or (read* N3 dysfluent) or (read* N3 dysfluenc*)
 S8 (read* N3 dysfunction*)
 S7 (read* N3 delay*)
 S6 (read* N3 defic*)
 S5 (read* N3 impair*)
 S4 (read* N3 abilit*) or (read* N3 disab*)
 S3 (read* N3 disorder*)
 S2 (MH "Reading Disorders+")
 S1 (MH "Reading+")

Science Citation Index - EXPANDED (SCI-EXPANDED), Social Science Citation Index (SSCI), Conference Proceedings Citation Index - Science (CPCI-S), Conference Proceedings Citation Index - Social Sciences & Humanities (CPCI-SSH); all Web of Science

#11 #10 AND #6 AND #1

#10 #9 OR #8 OR #7

9 TS=(“sight word*” or sight-word*)

8 TS=(lettersound* or letter-sound* or “letter sound*” or “letter identif*“)

7 TS=(phonics or phonem* or phonolog* or graphem*)

6 #5 OR #4 OR #3 OR #2

5 TS=(wordblind* or word-blind* or “word blind*“)

4 TS=(dyslexia or dyslexic*)

3 TS= (READ* SAME (accuracy or comprehen* or disorder* or disab* or abilit* or impair* or defc* or delay* or dysfunction* or dysfluen* or fluen*))

2 TS= (“slow read*” or “remedial read*” or “poor read*“)

1 TS=(random* or control* or trial* or group* or effectiveness or evaluation or placebo*)

ZETOC (zetoc.jisc.ac.uk)

Search terms: conference: reading phonics

ClinicalTrials.gov (ClinicalTrials.gov)

phonetics OR phonology OR phonics | reading OR dyslexia

World Health Organization International Clinical Trials Registry Platform (WHO ICTRP; www.who.int/ictpr)

CONDITION reading OR dyslexia

INTERVENTION : phonics OR phonetics OR phonology

metaRegister of Controlled Trials ISRCTN Registry (www.isrctn.com)

(reading or dyslexia) AND (phonics or phonology or phonetics)

ProQuest Dissertations and Theses Global

Search (key: AB (abstract); RTYPE (record type); SU (subject); TI (title); la (language))

((((AB (“randomly”) OR (RTYPE (“randomized controlled trial”) OR (RTYPE (“controlled clinical trial”) OR (AB (“randomi?ed”) OR (AB (placebo*)) OR (AB (“drug therapy”) OR (AB (“groups”) OR (AB (“trial”)) AND ((SU,exact(“reading”) OR TI,AB(read* NEAR/3 delay*) OR TI,AB(read*NEAR/3 disorder*) OR TI,AB(read NEAR/3 (ability OR disability)) OR TI,AB(read* NEAR/3 impair*) OR TI,AB(read* NEAR/3 defc*) OR SU,exact(“dyslexia”) OR TI,AB(read* NEAR/3 dysfunction*) OR TI,AB(poor* NEAR/3 read*) OR TI,AB(dysfluen* NEAR/3 read*) OR TI,AB(slow* NEAR/3 read*) OR TI,AB(remedial NEAR/3 read*) OR TI,AB(dyslex*) OR TI,AB(word-blind* OR word blind*)) AND ((TI,AB (sight word* OR sight-word*)) OR (TI,AB (phonics)) OR (TI,AB (phonem*)) OR (TI,AB (phonolog*)) OR (TI,AB (graphem*)) OR (TI,AB (lettersound* OR letter-sound*)) OR (TI,AB (letter identif*)) OR (SU, exact (“remedial teaching”)) OR (SU, exact(“phonetics”)) OR (TI,AB (read* NEAR/3 (teach* OR method* OR program*)))))) AND (((AB (“randomly”) OR (RTYPE (“randomized controlled trial”) OR (RTYPE (“controlled clinical trial”) OR (AB (“randomi?ed”) OR (AB (placebo*)) OR (AB (“drug therapy”) OR (AB (“groups”) OR (AB (“trial”)) NOT (AB (“exp animals/ not humans”)) AND la.exact(“ENG”))

DART Europe E-theses Portal (www.dart-europe.eu), Australasian Digital Theses program (adt.caul.edu.au), Australian Education Research Theses (www.acer.edu.au/library/theses), Networked Digital Library of Theses and Dissertations (NDLTD; www.ndltd.org), Theses Canada Portal (www.collectionscanada.gc.ca/thesescanada), www.dissertation.com, and www.thesisabstracts.com

Last searched July 2012. Replaced by ProQuest Dissertations and Theses Global for searches 2012 onwards.

1. dyslexia
2. reading disorder
3. reading disability
4. reading impairment
5. reading deficit
6. reading delay
7. reading dysfunction
8. poor reader
9. poor reading
10. dysfluent reader
11. dysfluent reading
12. slow reader
13. slow reading
14. remedial reader
15. word-blind
16. wordblind
17. phonics
18. phoneme
19. phonological
20. grapheme
21. lettersound or letter-sound
22. letter identification
23. sight word or sight-word
24. phonetics
25. remedial teaching
26. reading teaching
27. reading methods
28. reading program

WHAT'S NEW

Date	Event	Description
14 June 2018	New search has been performed	Review has been updated following a new search in February 2017 and May 2018
28 July 2017	New citation required but conclusions have not changed	Inclusion of 3 new studies has not changed main conclusions of review

HISTORY

Protocol first published: Issue 5, 2011

Review first published: Issue 12, 2012

Date	Event	Description
18 July 2017	New search has been performed	Summary of findings submission for update

CONTRIBUTIONS OF AUTHORS

All review authors were involved in designing the methodology; in extracting, analysing, and reporting data; and in checking and revising content of this review. As mentioned previously, review authors who were also authors on two included studies did not assess these studies for eligibility, extract data, or assess the risk of bias or the quality of the evidence.

The first author of the review, GMcA, is the guarantor.

DECLARATIONS OF INTEREST

GMcA:^{1,2} Director of the Macquarie University Reading Clinic (a non-profit organisation), and as such, she presents workshops to professionals about the treatment of reading difficulties. The money earned by such workshops goes to the clinic and GMcA declares that she does not benefit financially from these activities. Macquarie University covered GMcA's expenses to attend and present at various national and international conferences.

YS: none known.

NB: none known.

DF:¹ clinician (treatment) at the Macquarie University Reading Clinic.

HCW: Macquarie University covered her expenses to attend and present at various national and international conferences.

SK:^{1,2} Clinical Director of the Macquarie University Reading Clinic, and as such, she designs assessments and treatments, including those with a phonics component. In her role as Clinical Director, SK provides consultancy or professional development courses (or both) to parents, clinicians, schools, clinics, and the government. The money earned by these activities goes to the Macquarie University Reading Clinic and SK does not benefit financially from these activities. Macquarie University covered SK's expenses to attend and present at various national and international conferences. Between 2009 and 2010, SK was employed as a part-time postdoctoral researcher by MultiLit, a company which provides literacy instruction and sells literacy programs. These programs include a phonics component. SK was responsible for analysing and writing up data from students who received literacy instruction by MultiLit. SK does not receive financial benefits from the sale of any literacy programs.

EB:^{1,2} Clinic Co-Ordinator of the Macquarie University Reading Clinic.

TA:^{1,2} clinician (Assessment and Training) at the Macquarie University Reading Clinic.

EM: funded by the Australian Research Council as a Postdoctoral Research Fellow at the ARC Centre of Excellence for Cognition and its Disorders; the funds support research activities in general, and not specifically for doing this review. Macquarie University covered EM's expenses to attend and present at various national and international conferences.

AC:² none known.

¹Several authors on the revised version of the review (GMcA, SK, EB, TA, DF) work at the Macquarie University Reading Clinic, where they use phonics training for some poor readers (i.e. those with the appropriate profile), since the evidence suggests that this can be effective for some types of reading problems.

²Five review authors (GMcA, SK, AC, EB, TA) were involved in the conduct of two studies, which were included in this review update (McArthur 2015a; McArthur 2015b). None of these review authors assessed the eligibility of these studies for inclusion, extract data from these studies, or conducted the 'Risk of bias' and GRADE assessments.

Funds from the Australian Research Council, National Health Medical Research Council, Macquarie University Reading Clinic, and Macquarie University paid the wages of various authors during the development of the original review. These funds were provided for research activities in general, and not specifically for doing this review. For this review update, only HCW and EM received funds from ARC to support their wage, whereas all other review authors were supported by the Department of Cognitive Science at Macquarie University.

SOURCES OF SUPPORT

Internal sources

- Macquarie University, Australia.

Funds for the salaries of McArthur, Castles, Larsen, and Marinus

External sources

- National Health and Medical Research Council (NHMRC) Project Grant (488518), Australia.

Funds the salaries of Kohnen, Jones, and Banales

- Australian Research Council (ARC) Discovery Project Grant (DP0879556), Australia.

Funds for the salaries of McArthur, Anandakumar, and Larsen

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

1. In the Authors section, we replaced three authors (Pip Eve, Kristy Jones, and Linda Larsen) with three new review authors (YS, NS, and DF).

2. In the Review information section, we updated the name of one institution (the ARC Centre of Excellence of Cognition and its Disorders was previously called the Macquarie Centre for Cognitive Science).

3. [Description of the condition](#)

- i) We removed 'Figure 3' of the dual route model.

4. [Description of the intervention](#)

i) We provided a clearer definition of phonics training: "Phonics teaches people to read via phonics-based reading, which depends upon the abilities to: identify each letter or letter-cluster in a word (e.g. S H I P); transpose each letter or letter cluster into its correct speech sound ('sh' 'i' 'p') using the letter-sound rules; and blend these speech sounds into a word that can be said aloud ('ship')".

5. [Description of the intervention](#) and [How the intervention might work](#)

i) We provided a clearer explanation for why it is important to review simple phonics training programmes rather than complex programmes.

6. [Types of participants](#)

i) We clarified the inclusion criteria: "This review included studies that were conducted with poor readers who spoke English as their primary language at school or work, who lived in a country where English was the official language, and who were receiving phonics instruction in English."

7. [Types of outcome measures](#)

i) We renamed regular word accuracy as mixed/regular word reading accuracy as most studies that tested fluency using regular words included irregular words in the same test.

ii) We renamed regular word fluency as mixed/regular word reading fluency as all studies that tested fluency using regular words included irregular words in the same test.

iii) We merged 'spoken word production' and 'other phoneme awareness abilities' into 'phonological output' since these skills are tested with similar measures (i.e. phoneme awareness tests).

iv) We did not include one primary outcome (irregular word reading fluency) and four secondary outcomes (letter identification, parsing, blending, phoneme awareness) in either version of this review because no studies reported data for these measures.

v) Regarding timing of outcome assessment, all studies identified by this review reported data for outcomes immediately after training. Therefore, we had no data for the following time points: one to six months after training; seven to 18 months after training; or more than 18 months after training.

8. Electronic searches

i) On the advice of our Cochrane Information Specialist, we replaced a series of 'free' but unproductive sources (DART Europe E-theses Portal, Australasian Digital Theses Program, Education Research Theses, Electronic Theses Online Service, Networked Digital Library of Theses and Dissertations; Theses Canada portal, www.dissertation.com, and www.thesisabstracts.com), with ProQuest Dissertations and Theses Global.

ii) On the advice of our Cochrane Information Specialist, we searched the ISRCTN registry because the metaRegister of Controlled Trials is under review.

iii) On the advice of our university librarian, when we searched ProQuest Dissertations and Theses Global, we only included studies that were published in English since our focus was English-speaking poor readers.

9. Data synthesis

i) In our protocol, [McArthur 2011](#), we planned to synthesise similar types of poor readers (mixed, phonological, surface, unknown). However, the studies included in this review predominantly had mixed poor reading rather than phonological or surface dyslexia.

10. Subgroup analysis and investigation of heterogeneity

i) As with the previous version of this review ([McArthur 2012](#)), we did not conduct subgroup analyses of "poor-reading profile" and "spoken language" because no study provided relevant data to divide the studies into appropriate subgroups.

ii) We chose not to report the results of our five subgroup analyses because no subgroup included more than nine studies (most only comprised two to seven studies), and the heterogeneity of data with most subgroups (particularly the larger ones) was high (i.e. I^2 greater than 70; see Subgroup analyses under [Effects of interventions](#)).

NOTES

None.

INDEX TERMS

Medical Subject Headings (MeSH)

*Reading; Dyslexia [*rehabilitation]; Language; Phonation [*physiology]; Randomized Controlled Trials as Topic

MeSH check words

Adolescent; Adult; Child; Female; Humans; Male