the british psychological society
promoting excellence in psychology

British Journal of Educational Psychology (2021)
www.wileyonlinelibrary.com

Decoding and comprehension skills mediate the link between a small-group reading programme and English national literacy assessments

Janet I. Vousden[1]*, Anna J. Cunningham[1], Helen Johnson[1], Sam Waldron[1], Sabrina Ammi[1], Claire Pillinger[1], Robert Savage[2] and Clare Wood[1]

[1]Coventry University, UK
[2]McGill University, Montreal, Quebec, Canada

Background. Despite the fact that literacy instruction is a main focus of primary education, many children struggle to meet nationally set standards.

Aims. We aimed to test which components of a comprehensive reading programme (ABRACADABRA: https://eur03.safelinks.protection.outlook.com/?url=https%3A%2F%2Fdoi.org%2F10.1186%2FISRCTN18254678&data=04%7C01%7Cjanet.vousden%40ntu.ac.uk%7C880280e0b00749df855308d94068a0bb%7C8acbc2c5c8ed42c78169ba438a0dbe2f%7C1%7C0%7C637611640381216902%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C1000&sdata=%2B4U9sGfofkyCPEY7IWz8n3TPoMOAeJMXyFwdhW6EpUw%3D&reserved=0) mediated the effect of the programme on nationally assessed literacy outcomes.

Sample. Following blind allocation, 516 Year 1 pupils from 40 schools were randomized to the programme group, and 908 Year 1 pupils, to a control condition.

Methods. Pupils in the programme completed 20 weeks of instruction in grapheme/-phoneme knowledge, decoding, and comprehension. Control children received regular classroom instruction.

Results. Children in the programme group were significantly better at these taught skills after the programme finished (effect sizes: grapheme/phoneme knowledge, $\beta = .33$, 95% CI [0.09–0.57]; decoding, $\beta = .26$, 95% CI [0.09–0.43]; and comprehension, $\beta = .26$, 95% CI [0.05–0.47]). Improvements in the programme group's decoding and comprehension skills fully mediated the improvements in national literacy assessments serving as a delayed post-test 12 months after the programme. Programme group pupils were 2.3 (95% CI [1.4–4.1]) times more likely to achieve/exceed the expected standard in reading, and 1.8 (95% CI [1.2–2.6]) times more likely to achieve/exceed the expected standard in writing due to an increase in the trained skills.

*Correspondence should be addressed to Janet Vousden, Department of Psychology, Nottingham Trent University, 50 Shakespeare Street, NG1 4FQ, UK (email: janet.vousden@ntu.ac.uk).

DOI:10.1111/bjep.12441

Conclusions. These results provide strong evidence that a programme that incorporates decoding and comprehension instruction for typically developing beginning readers improves distal educational outcomes in reading and writing through increasing proficiencies targeted by the reading programme.

One of the main objectives for primary or elementary school education is pupils being able to read and write to a level that will enable them to access other areas of the curriculum within secondary education and beyond. Globally, there is still much work to be done. In the 2018 Organisation for Economic Co-operation and Development (OECD) Programme for International Student Assessment (PISA) results, 34 countries showed no improvement or a decline in reading standards over a three-year period compared with only 15 countries that showed an improvement; almost one in four children in the OECD counties were unable to read proficiently for learning (Schleicher, 2019). In England, where the present work is situated, of the subjects assessed after the first three years in primary education (known as Key Stage 1; reading, writing, science, and mathematics), reading and writing have the smallest proportion of children performing at or above the expected standard. Data from 2019 showed that 25% and 31% of children did not achieve the expected standard in teacher-assessed reading and writing at Key Stage 1, respectively (Department for Education, 2019). One way in which literacy outcomes could be improved is through effective interventions at an early age. One such programme is the Year 1 reading programme, which has been shown to be effective at improving Key Stage 1 outcomes for reading and writing (Johnson et al., 2019). In the current paper, we assess the reasons behind the effectiveness of this programme by investigating the skills, which mediated the link between participation in the programme and Key Stage 1 outcomes. In doing so, we establish the roles of decoding and comprehension to explain causal links between instruction and outcomes.

Much of the focus in early literacy in England is on decoding skills (Rose, 2006); children need these skills to be able to read words aloud and spell them. Less emphasis is placed on text comprehension at this point, even though higher level skills and knowledge such as structural knowledge and inferencing form part of Key Stage reading assessments (Department for Education, 2018). This paper presents the results of the Year 1 reading programme, which teaches primary aged children (aged 5–6 years) reading comprehension skills and decoding level skills. Our aim is to determine whether the previously observed positive effects of the reading programme on Key Stage 1 reading and writing assessments can be reliably attributed to an increase in both decoding and comprehension abilities, as targeted by the programme.

Research has shown phonics to be an effective method for most children to learn to read (Castles, Rastle, & Nation, 2018; Ehri, 2020; Ehri, Nunes, Stahl, & Willows, 2001). Phonics teaching focuses on the systematic relationships that exist between print and sound (letter-sound knowledge, LSK) and the associated oral skills necessary to turn print into sound (phonological awareness, PA). PA is strongly associated with reading, with evidence suggesting a causal role in predicting later reading ability (Hulme, Bowyer-Crane, Carroll, Duff, & Snowling, 2012; Melby-Lervåg, Lyster, & Hulme, 2012; Muter, Hulme, Snowling, & Stevenson, 2004). Early knowledge of letter sounds and names is also a strong predictor of later reading ability (Hulme & Snowling, 2013; Schatschneider, Fletcher, Francis, Carlson, & Foorman, 2004). Phonics-based programmes have been shown to be effective for teaching reading (Ehri, 2020; Ehri et al., 2001; Stuart, 1999; Torgerson, Brooks, Gascoine, & Higgins, 2019) although the effects of training skills in isolation can be small (Piasta & Wagner, 2010). PA and LSK seem inextricably linked

(Castles & Coltheart, 2004), with Byrne and Fielding-Barnsley (1989, 1991) finding evidence that both are needed for children to read, and Hatcher, Hulme, and Snowling (2004) showing that PA training is most effective when combined with LSK training (see also Savage et al., 2020; Yeung & Savage, 2020). In summary, for decoding at least, the evidence suggests it is important to train both components within the same programme.

However, it is important both theoretically and from a pedagogical perspective, to demonstrate that the effectiveness of a reading programme is due to gains in the component skills and knowledge delivered within the programme rather than other potential sources of influence. RCTs of interventions coupled with mediation models are amongst the strongest evidence of causality currently available in the social sciences. For example, Hulme et al. (2012) used a mediation model to demonstrate the roles played by the components of a reading intervention for primary school children with weak oral language skills (Bowyer-Crane et al., 2008). Hulme et al. (2012) showed that an increase in a combined reading and spelling literacy outcome was mediated by gains in LSK and PA skills taught to the intervention group. This is a useful approach for disentangling the likely pathways through which instructional programmes achieve their outcomes, especially when programmes contain multiple components. Additionally, this is important because it demonstrates that the impact achieved by the programme is related to the programme content, rather than due to other untested variables.

While much research has been conducted on phonics in early reading, there are fewer reading comprehension interventions for early readers (Castles et al., 2018). Reading comprehension is a highly complex process that requires a certain level of decoding ability to be in place first, and this is not always demonstrable in early readers. The literature shows that comprehension is a complex cognitive process, best facilitated through a multitude of approaches (Language and Reading Research Consortium , Logan (LARRC), 2019; National Institute of Child Health & Human Development, 2000) both at single-word and at textual levels. At the single-word level, vocabulary is strongly linked to comprehension, with vocabulary instruction often yielding positive effects on reading comprehension (e.g., Elleman, Lindo, Morphy, & Compton, 2009). At a textual level, a range of diverse skills such as comprehension monitoring (e.g., Justice et al., 2017; Oakhill & Cain, 2012), collaborative learning (e.g., Clarke, Snowling, Truelove, & Hulme, 2010), and story structure (e.g., Cain, Oakhill, & Bryant, 2004; Oakhill & Cain, 2012) all help to improve reading comprehension. Outcomes are more likely to be positive if these techniques are combined rather than taught in isolation (National Institute of Child Health & Human Development, 2000). Therefore, one option is is to teach early readers a range of comprehension techniques in combination with decoding skills within the same intervention. This was the case for the Year 1 reading programme, which taught the two skills in parallel. However, it is not yet known whether the reason why it was successful was due to improvements in decoding and comprehension skills specifically.

Comprehension often incorporates higher-order skills concerned with structural aspects of text (e.g., identifying key parts of a text for summarizing). Writing also draws on structural aspects of texts due to its compositional nature (Berninger et al., 2002; Kim & Schatschneider, 2017). It is thus not surprising then that some skills associated with reading comprehension such as comprehension monitoring (Limpo, Alves, & Fidalgo, 2014), summarizing (Mason, Snyder, Sukhram, & Kedem, 2006), vocabulary (Savage, Kozakewich, Genesee, Erdos, & Haigh, 2017), and story structure (Arrimada, Torrance, & Fidalgo, 2019; Pinto, Tarchi, & Bigozzi, 2016; Spencer & Petersen, 2018) also benefit writing. Increasing reading comprehension ability through instruction in these skills therefore should also confer benefits on writing ability.

Teaching comprehension and decoding skills together are important for early readers, as both have been shown to be predictive of reading ability from a young age. For example, Whitehurst and Lonigan (1998) suggested two interdependent sets of skills are involved in conventional literacy ability: both 'inside-out' skills, which broadly encompass the requisite skills for decoding, including PA and LSK skills; and 'outside-in' skills, which relate to understanding the context of written text, including narrative understanding and conceptual skills. Although the inside-out skills were stronger predictors of early reading, outside-in skills such as vocabulary clearly contributed both directly and indirectly to reading in the first and second grade as well (Whitehurst & Lonigan, 1998).

In line with the influential simple view of reading (Gough & Tunmer, 1986), recent evidence suggests that reading comprehension is dependent on both decoding and oral language skills (Lervåg, Hulme, & Melby-Lervåg, 2018; Lonigan, Burgess, & Schatschneider, 2018). Consistent with this, improving reading comprehension in young readers can be achieved through improving oral language (e.g., Oakhill, Cain, & Bryant, 2003). For example, multicomponent oral language interventions have been successful at boosting reading comprehension in children aged 7–9 years (Language and Reading Research Consortium (LARRC) et al., 2019; Williams, Brooke, Lauer, Hall, & Pollini, 2009; Williams et al., 2014). However, despite theoretical reasons to introduce comprehension activities from the start of reading instruction, few high-quality RCT interventions have focused on beginning readers (i.e., younger than age 7; see Rogde, Hagen, Melby-Lervåg, & Lervåg, 2019, for a recent meta-analytic review). Where researchers have done so they have not involved the training of large numbers of regular Year 1 (roughly equivalent to grade 1) classroom teachers, and are often limited to intervention work for children with poor oral language or literacy skills (Bowyer-Crane et al., 2008; Clarke et al., 2010; Edmonds et al., 2009; Scammacca, Roberts, Vaughn, & Stuebing, 2015).

The Year 1 reading programme drew on reading activities from ABRACADABRA (A Balanced Reading Approach for Canadians Designed to Achieve Best Results for All, ABRA; Abrami et al., 2010), which has been trialled internationally with consistently positive results (Abrami, Lysenko, & Borokhovski, 2020; McNally et al., 2016; Bailey, Arciuli, & Stancliffe, 2017; Savage, Abrami, Abrami, Hipps, & Deault, 2009; Savage et al., 2010). These include both small-scale trials on beginning readers with a trained student research assistant (RA) teacher (Savage, Abrami, et al., 2009), and larger trials with trained teachers instead of student RAs (Savage et al., 2013). While studies such as these are promising demonstrations of improvements in both phonic and comprehension skills within the same intervention, it is not clear whether or how these skills might contribute to distal educational literacy outcomes, nor whether they scale up with trained wider school staff, such as teaching assistants. The present study thus represents progression from demonstrations of internal validity of programmes to external validity via scaled-up studies with trained school staff, educationally important outcomes, and robust intention to treat designs. In the present study, we combine an RCT of the Year 1 reading programme with a mediation analysis, to assess which parts of the programme affect performance in Key Stage 1 national literacy assessments in reading and writing. This approach provides strong support for causal explanations of trained skills and a possible model for future scale-up.

The current study

The aim of the current study was to understand which of the taught skills and knowledge in the Year 1 reading programme (McNally et al., 2016; Johnson et al., 2019) mediated the

effects on pupils' teacher-assessed Key Stage 1 literacy outcomes a year after the programme finished. Children in the programme group were found to have better outcomes for reading and writing, but it is not clear to what extent this result is attributable to improvements in skills and knowledge taught within the programme: grapheme–phoneme knowledge, decoding, comprehension, or a combination of all three. We present a new analysis of the data from the original study of McNally et al. (2016), using structural mediation models to examine the indirect and direct effects of the reading programme and its component parts on Key Stage 1 outcomes.

Method

Design

This study is a pre-test–post-test RCT design that took place over one academic year and proceeded according to an intention-to-treat protocol that was agreed before the study began (registered on the ISRCTN website, https://doi.org/10.1186/ISRCTN18254678).

Randomization of pupils was a two-stage process. First, schools were matched on school administrative data (size of the relevant year group, average point score from teacher assessments of children aged 7 in school,[1] and the percentage of pupils eligible for free school meals) as this increases the power of the RCT (Spybrook et al., 2011). Matched schools were randomly allocated to either the treatment or the control group. Pupils in the control group of schools continued with their regular classroom teaching. Second, pupils in the treated group of schools were randomly allocated to one of three groups: (1) control, (2) Web-based programme, and (3) non-Web-based programme, although the Web and non-Web-based programmes were collapsed for this study. For the purpose of this study, the control children in treated schools were excluded from the analysis because the original evaluation showed spillover effects (McNally et al., 2016). However, there were no spillover effects at delayed post-test in the original study, and thus, there are no threats to internal validity in the current design. Our two comparison groups for this study were pupils in control schools and children in either programme group.

Participants

Schools

All eligible non-selective primary schools in the Midlands region in the United Kingdom were invited to participate in the study. Fifty schools were recruited under these conditions.

Five schools assigned to the treatment group dropped out. As we were interested in the mechanisms in the programme responsible for the outcomes, we excluded any school that dropped out and did not complete the programme.[2] Both the dropped-out schools and their randomization pair were excluded from the analysis (N = 10). This left 40 schools for analysis. School characteristics in terms of staffing and pupil intake are described in the Supporting Information. There were no differences between the forty schools retained for analysis and the 10 schools excluded from analysis in this respect.

[1] *This was based on the average point score at the end of the first Key Stage (1) at age 7.*

[2] *Three schools dropped out immediately after randomization, one dropped out less than 5 weeks into the programme, and one school dropped out less than halfway through the programme.*

Pupils

All children in the relevant year group from signed-up schools were selected for participation, regardless of ability, unless the pupil's guardian withdrew them from the study. Of the 1,721 children in the 40 schools, 282 were in the control group in the treated schools and were removed from analysis (see earlier), leaving 1,439 children ($N = 520$ in the programme group and $N = 919$ in the control group) across the 40 schools. The average age of participating pupils was 5 years and 6 months ($N = 733$ female and $N = 706$ male).

Attrition. Pupils were assessed at three time points: Time 1 (T1, pre-test, at the beginning of Year 1 in the autumn term); Time 2 (T2, post-test; within one month of the programme end, at the end of Year 1 in the summer term); and Time 3 (T3, delayed post-test; one year after the programme had concluded, at the end of Year 2 in the summer term). Of the sample of 1,439 children tested at T1, 1380 were also tested at T2 (T1-T2 attrition of 4.1%), and 1,364 were tested at T3 as well, and had Key Stage 1 data available at T3 (T2-T3 attrition of 1.2%). The main reason for attrition between T1 and T2 was moving school. Some Key Stage 1 data were unavailable from the National Pupil Database (NPD) at T3, accounting for the T2-T3 attrition.

Procedure

The Year 1 reading programme

Content. The programme was developed to supplement the national curriculum in literacy in England (UK) for children in their second year of formal schooling (Year 1). It consisted of a comprehensive range of short activities to support decoding (both sublexical grapheme–phoneme knowledge and word reading), fluency, and comprehension, practised alongside real stories (e.g., Aesops Fables), and is summarized in the Supporting Information.

Delivery. The programme was delivered four times a week over a 20-week period from November to May of Year 1. Each session lasted 15 min and was conducted in small groups of four pupils, totalling 20 h of programme time. Sessions did not replace the core literacy provision of the school and instead were conducted during lessons in which pupils would normally be doing something related to literacy, for example, topic work (where pupils learn through reading and writing about topics linked to the curriculum, such as history through heroes and heroines) or guided reading. This balanced the total amount of literacy instruction received by treatment and control groups. Sessions were led by a trained teaching assistant (TA). Details of TA characteristics can be found in the Supporting Information.

Two versions of the programme were delivered: A Web-based version was delivered on laptops using ABRA software (Abrami et al., 2010), and a non-Web-based version was delivered with paper materials. The versions were closely matched, and the same lesson plans were used.[3] The original evaluation showed that both delivery methods were

[3] *Vocabulary was slightly different. Web-based groups read two example sentences for a new vocabulary word and then did a two-way forced-choice test on a further two example sentences, one right and one wrong. The non-Web-based vocabulary activity asked children to suggest words relating to the new vocabulary word (creating a semantic network).*

associated with a positive effect for Key Stage 1 reading and writing, with no significant difference between them (McNally et al., 2016; Johnson et al., 2019). We therefore collapsed across delivery method for this study, to evaluate the effect of the pedagogy of the programme itself, rather than the delivery medium, against a control group.

Training. Teaching assistants (TAs) already in post at participating schools attended one and a half day's training with the programme delivery team, where they were trained how to deliver the programme. Once the programme had started, TAs received 'just-in-time' support from the delivery team via email or phone as and when they requested help throughout the duration of the programme. Training was evaluated by an independent evaluation team, and the TAs received two on-site visits from the delivery team to assess treatment integrity. Further details about training and treatment integrity measures can be found in the Supporting Information.

Measures

Reliability data are presented in Table 1.

Early Years Foundation Stage Profile (Time 1)

General ability at baseline was measured by the Early Years Foundation Stage Profile Good Level of Development (FSP_GLD). The profile is teacher-assessed at the end of the first year of formal schooling (at age 5) and measures attainment against expected levels in five areas: communication and language, physical development, personal social and emotional development, literacy, and mathematical development (Standards & Testing Agency, 2013). The FSP_GLD variable reported by schools is dichotomous, indicating whether pupils have 0 (not achieved the expected level) or 1 (achieved or exceeded the expected level) in all five key areas of development (see the Supporting Information for further information).

Disadvantage (Time 1)

Pupil disadvantage was measured according to whether participating pupils were eligible for free school meals (FSM) in school on a census day in the school year, and this measure was used as a covariate. FSM data were obtained from the NPD.

Grapheme- and phoneme-level knowledge (Time 1 and Time 2)

Phoneme-level knowledge was measured by the phoneme segmentation subtask of the Pre-Reading Inventory of Phonological Awareness (PIPA; Dodd, Crosbie, McIntosh, Teizel, & Ozanne, 2000). Children were instructed to use plastic counters to indicate the number of phonemes in 12 spoken words. Robust procedures and data for content, concurrent, criterion-related, and construct validity are reported in the PIPA manual. Letter-sound knowledge was measured using the Letter-Sound Test (LeST, Larsen, Kohnen, Nickels, & McArthur, 2015) at Time 1 and Time 2. Graphemes were presented 13 to a sheet, and children were asked to provide the sound for each grapheme. There were 25 single letters and 26 digraphs in total. There was no stopping rule. Good criterion validity is reported for LeST (Larsen et al., 2015).

Table 1. Means and standard deviations of measures taken at T1, T2, and T3

Variable	Max	*N* at T1	Time 1 Programme group *M* (*SD*)	*N* at T1	Time 1 Controls *M* (*SD*)	*N* at T2	Time 2 Programme group *M* (*SD*)	*N* at T2	Time 2 Controls *M* (*SD*)	*N* at T3	Time 3 Programme group *M* (*SD*)	N at T3	Time 3 Controls *M* (*SD*)	Reliability
Grapheme/phoneme knowledge														
Phoneme segmentation	12	512	5.94 (3.66)	894	5.88 (3.66)	487	8.37 (3.10)	872	8.08 (3.14)	–	–	–	–	0.70[a]
Letter-sound knowledge	51	512	33.18 (8.24)	883	34.57 (8.14)	487	44.91 (5.67)	871	43.85 (6.07)	–	–	–	–	0.88[b]
Decoding														
Non-word reading	30	510	5.35 (4.65)	857	5.55 (5.05)	485	14.46 (7.11)	870	13.27 (7.18)	–	–	–	–	0.96[a]
Regular word reading	30	510	6.40 (5.98)	856	6.91 (6.64)	485	17.53 (7.42)	870	16.69 (7.86)	–	–	–	–	0.97[a]
Exception word reading	30	510	3.34 (4.58)	857	3.80 (5.09)	485	11.46 (6.94)	869	10.92 (7.18)	–	–	–	–	0.97[a]
Comprehension														
Vocabulary	–	505	69.78 (16.66)	868	73.34 (15.88)	483	81.10 (14.77)	864	85.19 (14.66)	–	–	–	–	0.91[a]
PIRA: literal	8	442	2.42 (2.41)	792	2.56 (2.41)	484	6.63 (3.77)	869	6.10 (3.99)	–	–	–	–	0.93[c] (T1) 0.94[c] (T2)
PIRA: meaning	5 (T1) 13 (T2)	376	1.62 (1.42)	694	1.80 (1.49)	435	3.65 (2.05)	742	3.48 (2.16)	–	–	–	–	
Controls														
FSM (per cent eligible)	–	516	22.5	908	20.0	–	–	–	–	–	–	–	–	–
FSP_GLD (per cent achieve)	–	516	44.2	908	56.1	–	–	–	–	–	–	–	–	–[d]
Outcomes														
KS1 reading (per cent achieve)	–	–	–	–	–	–	–	–	–	513	72.5	891	71.5	–[d]
KS1 writing (per cent achieve)	–	–	–	–	–	–	–	–	–	513	62.0	891	59.0	–[d]

Note. Without missing data; controls, $n = 919$, programme group, $n = 520$. Per cent achieve means percentage of pupils who achieved expected or exceeded for that assessment. Reliabilities are published, taken from the test manuals.

[a]Cronbach's alpha.; [b]Intra-class correlation.; [c]Published reliability for whole PiRA.; [d]Although published reliability and validity measures are not publicly available from the Standards and Testing Agency (to the best of our knowledge), the KS1 teacher assessments must take into account results of the statutory KS1 tests. There is also evidence from published works that EYFSP and KS1 teacher assessments correlate with other standardized measures (Juniper Education, 2021; McCarty & Ruttle, 2010; Snowling, Hulme, Bailey, Stothard, & Lindsay, 2011), behave similar to standardized tests in terms of the baseline measures that predict them (Savage & Carless, 2004, 2005), and are sensitive to responsiveness to intervention (Savage & Carless, 2008; Savage, Carless, Carless, & Erten, 2009). Additionally, FSP_GD and the Key Stage 1 outcomes in this study correlate in expected ways with the other standardized tests. Together, this increases confidence that the teacher assessments are reliable and valid measures.

Decoding (Time 1 and Time 2)

Single-word reading was measured using the Diagnostic Test of Word Reading (DTWRP; Forum for Research in Literacy & Langauge, 2012). Children were asked to read aloud words presented on a sheet. Children read from three subtests, which included 30 words each; regular words, exception words, and pseudowords. Reading from each subtest was stopped after five consecutive errors. Robust procedures for concurrent validity are reported in the DTWRP manual.

Vocabulary (Time 1 and Time 2)

The British Picture Vocabulary Test (BPVS 3; Dunn & Dunn, 2009) is a test of receptive vocabulary and was administered individually to pupils. Pupils were asked to indicate which of four pictures best matched a spoken word. A robust validation process is reported in the BPVS manual.

Comprehension (Time 1 and Time 2)

Comprehension was measured using the comprehension subcomponents of the Progress in Reading Assessment (PiRA; McCarty & Ruttle, 2010). PiRA measures decoding and comprehension skills with three subcomponents – decoding (including phonics), literal comprehension, and reading for meaning (inference and prediction). There are three different PiRA test papers for each year group (one per term), designed to be approximately matched to the national curriculum. Pupils took the Autumn Term Test at Time 1 and the Summer Term Test at Time 2. Pupils were assessed in groups of 3–4 under the supervision of an RA. RAs instructed each group throughout the test, but pupils completed the test booklets on their own. In the present study, we use only the results of the literal and inferential components of the PiRA. Criteria for reporting validity are presented in the PiRA manual.

English national assessments in reading and writing (Key Stage 1 literacy assessments, Time 3)

At the end of Key Stage 1 (the third year of formal schooling when the child turns 7), teachers submit a judgement on the reading and writing attainment of each child in their class. These were our outcome variables in this study and were collected one year after the programme. In the current study, the data were dichotomized such that children were either 0 (working towards) or 1 (working at or working at greater depth), the expected standard. This was done in order to reflect the greater educational relevance of predicting which children and why did not achieve/achieved a minimum standard versus qualifying this standard into multiple levels.

The Key Stage 1 teacher assessments are administered according to a standardized framework of 'pupil can' statements (Standards & Testing Agency, 2016; see the Supporting Information for further information).

All measures (except the end of key stage and FSM measures) were collected by trained RAs that were blind to the nature of the study and the school and pupil allocations. Measures were administered according to the standardized assessment instructions. Key Stage 1 reading and writing assessment data, as well as FSP_GLD and FSM status data, were obtained from the NPD. Additional measures from the original study that were unrelated to the present research aims are not presented here.

Analysis plan

SEMs with mediation were built in Mplus version 8.3 (Muthen & Muthen, 2017), allowing for simultaneous testing of direct and (multiple) indirect effects. SEM is a technique frequently used for mediation analysis (Arias et al., 2016; Arlinghaus, Lombardi, Willetts, Folkard, & Christiani, 2012), preferred for its superior power to multistep regression methods (Iacobucci, Saldanha, & Deng, 2007). As the Key Stage 1 literacy assessments we used as outcomes were dichotomous variables, we used logistic regression analysis with MLR (maximum-likelihood estimation with robust standard errors) and exponentiated coefficients to yield odds ratios. Consistent with Authors and Authors (refs), to allow for the non-independence of observations arising from the clustering of children within schools, we used the robust Huber–White cluster-corrected standard errors in all models. Missing data were handled by full-information maximum-likelihood estimation.

Latent variables in the SEM reflected the main components of the skills and knowledge taught in the programme: decoding and comprehension. Decoding was represented by two latent variables to reflect sublexical skills and knowledge on the one hand (grapheme/phoneme knowledge), and word-level decoding (decoding) on the other. Comprehension was represented by one latent variable (comprehension). These latent constructs reflect the 'simple view of reading', which postulates that reading is the product of decoding and comprehension (Gough & Tunmer, 1986). As continuous variables based on experimenter-administered tests, they are distinct from the teacher-assessed literacy levels designated at the end of Key Stage 1.

Two mediation models were built using SEM, to understand how the component skills and knowledge taught in the reading programme affected two more distal educational assessments of literacy (Key Stage 1 reading and Key Stage 1 writing).

Results

Data preparation

Of the 1,439 children who were tested at T1, 15 had at least one missing data point for a binary covariate (FSP_GLD and FSM) and were not included in analyses.[4] This resulted in a final sample of 1,424 children for analyses (516 in the programme group and 908 in the control group). The mean age for both the programme group and control group at T1 was 5 years 6 months. The programme group was 47.2% male, and the control group was 49.8% male. Across the data set, 5.6% of data were missing (see the Supporting Information for missing data analysis).

Table 1 shows descriptive statistics, *N*, and published reliabilities for all measures collected, plus frequencies for NPD data, split by group (programme and controls). Reliabilities were consistently high (0.83–0.97). Table 2 shows the correlations between all observed variables used in our models. All correlations were significant, except some with phoneme segmentation (likely due to ceiling effects in this variable). There was no notable difference in the pattern of correlations between the programme group and the control group.

[4] *Analyses on the sample of 1,439 children using imputed values for missing binary covariate data (imputed using fully conditional specification) resulted in the same pattern of results.*

Table 2. Correlations between observed variables

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1. Phoneme segmentation T1	–	.56	.49	.46	.33	.31	.40	.33	.36	.44	.53	.56	.47	.30	.44	.35	−.15	.40	.47	.43
2. Letter-sound knowledge T1	.53	–	.71	.73	.64	.39	.55	.56	.26	.66	.66	.70	.72	.41	.61	.51	−.22	.47	.53	.49
3. Non-word reading T1	.41	.66	–	.87	.80	.41	.55	.58	.17	.42	.60	.62	.67	.44	.53	.43	−.16	.43	.44	.44
4. Regular word reading T1	.41	.69	.85	–	.89	.41	.60	.62	.11	.44	.61	.65	.73	.45	.57	.47	−.15	.47	.44	.45
5. Exception word reading T1	.31	.63	.80	.88	–	.31	.55	.62	*.05*	.35	.50	.54	.68	.37	.50	.42	−.10	.41	.33	.35
6. Vocabulary T1	.29	.37	.30	.32	.28	–	.41	.33	.13	.21	.28	.34	.35	.79	.42	.30	−.15	.38	.41	.40
7. PIRA: literal T1	.38	.57	.52	.58	.57	.30	–	.73	*.03*	.35	.42	.49	.53	.49	.52	.39	−.19	.38	.38	.41
8. PIRA: meaning T1	.35	.57	.53	.61	.58	.33	.76	–	*.02*	.35	.44	.50	.56	.40	.50	.45	−.19	.32	.32	.34
9. Phoneme segmentation T2	.27	.22	.11	.13	*.07*	.24	*.01*	*.00*	–	.39	.29	.27	.16	.11	.24	.18	*−.07*	*.06*	.24	.17
10. Letter-sound knowledge T2	.45	.68	.47	.49	.43	.35	.43	.43	.36	–	.65	.70	.63	.26	.59	.47	−.18	.32	.54	.49
11. Non-word reading T2	.48	.68	.66	.69	.63	.28	.52	.53	.23	.66	–	.86	.75	.30	.68	.53	−.14	.38	.58	.55
12. Regular word reading T2	.53	.72	.64	.67	.60	.36	.55	.57	.26	.73	.87	–	.83	.39	.73	.60	−.16	.46	.68	.65
13. Exception word reading T2	.46	.74	.69	.76	.74	.36	.62	.64	.13	.67	.82	.86	–	.40	.74	.61	−.17	.44	.60	.61
14. Vocabulary T2	.30	.38	.34	.36	.32	.74	.38	.40	.19	.36	.35	.42	.41	–	.43	0.32	−0.17	0.36	0.40	.37
15. PIRA: literal T2	.48	.65	.57	.61	.57	.43	.57	.59	.16	.62	.68	.76	.76	.49	–	.73	−.14	.46	.59	.61
16. PIRA: meaning T2	.41	.56	.47	.56	.51	.35	.45	.48	*.07*	.53	.60	.69	.70	.41	.78	–	*−.05*	.39	.45	.49
17. FSM	−.12	−.16	−.11	−.13	−.12	−.20	−.14	−.15	*−.07*	−.19	−.14	−.17	−.18	−.17	−.21	−.17	–	−.11	−.18	−.14
18. FSP_GLD	.47	.50	.41	.43	.37	.42	.44	.44	.15	.47	.48	.54	.51	.41	.55	.46	−.13	–	.40	.45
19. KS1 Reading	.46	.57	.41	.43	.38	.33	.42	.43	.20	.59	.59	.69	.62	.35	.62	.54	−.15	.48	–	.71
20. KS1 Writing	.38	.52	.44	.46	.41	.30	.46	.48	.19	.52	.57	.62	.60	.32	.59	.47	−.13	.53	.70	–

Note. Correlations between continuous variables are bivariate (Pearson's *r*). All values are significant ($p < .05$) unless in italics. Correlations between a binary and continuous variable are point-biserial, and correlations between two binary variables are phi. Children in the programme group are shown above the diagonal, and control group children are shown below the diagonal.

Confirmatory factor analysis

Confirmatory factor analysis was used to confirm the structure of our latent variables (grapheme/phoneme knowledge, decoding, and comprehension) at T1 and T2, using standardized raw scores for each of the indicator variables. Grapheme/phoneme knowledge was indicated by phoneme segmentation (PIPA) and letter-sound knowledge (LeST); decoding was indicated by total correct for the three DTWRP tasks (non-word, regular word, and exception word reading); and comprehension was indicated by total correct for literal questions in the PiRA, total correct for meaning questions in the PiRA, and BPVS vocabulary. Both models showed an excellent fit to the data; T1: CFI = .98, RMSEA = .067 (90% CI: 0.056–0.078), SRMR = .027 (Figure 1); and T2: CFI = .98, RMSEA = .070 (90% CI: 0.059–0.080), SRMR = .024 (Figure 2).

Mediation models

We conducted the mediation analyses following the approach outlined in Hayes (2018) by testing the full model with mediation, rather than to establish more traditional criteria for mediation, because of recent consensus that the establishment of a total effect should not be a prerequisite for testing indirect effects (McKinnon, 2008; Zhao, Lynch, & Chen, 2010; for a full discussion, see Hayes, 2018) and that tests of indirect effects can have more power than tests of total effects (Kenny & Judd, 2014). This model included links from

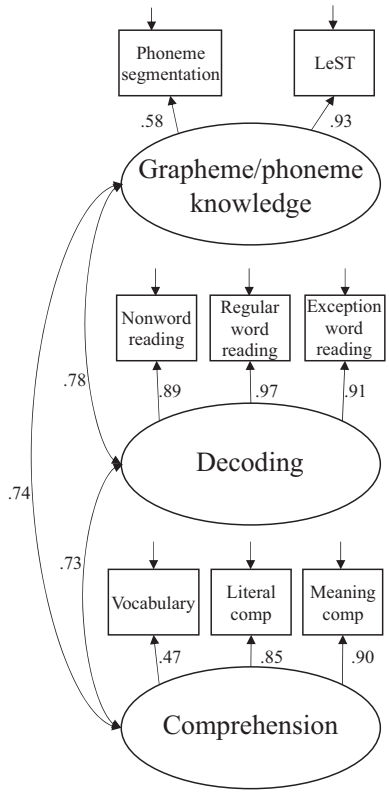

Figure 1. Confirmatory factor analyses for latent variables at Time 1. Standardized factor loadings are given next to each link from latent variable to indicator variable. Covariances are indicated by curved arrows with r values next to them.

programme to the mediators (G/P knowledge T2, decoding T2, and comprehension T2) and then from the mediators to the outcome (Key Stage 1 reading or writing). In addition, a direct link was included from programme to outcome. Baseline scores for FSP_GLD and FSM were covaried at Time 1, as well as controlled for through direct links with the mediators and outcome (to control for initial differences in ability and socioeconomic status). Also, error variances for latent variables at Time 2 were covaried, to allow for high correlations between ancillary task demands for the tests of different constructs (e.g., ability to follow instructions). Finally, the autoregressors (G/P knowledge, decoding, and comprehension at T1) were regressed onto their equivalents at T2. Unfortunately, none of the 'traditional' model fit indicators (such as the chi-square) can be applied to these specific models here, which include logistic regression analysis with full-information maximum-likelihood estimation (FIML). However, the excellent fit of the CFAs, together with the strong pattern of correlations reported in Table 2, leads us to be confident in our mediation models.

Calculation of indirect effects

Mplus (version 8.3; Muthen & Muthen, 2017) multiplies together the non-standardized regression coefficients from predictor to mediator, and mediator to outcome, to get the

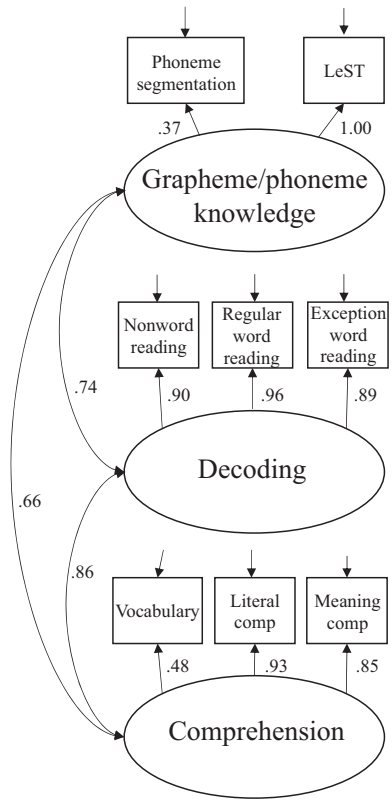

Figure 2. Confirmatory factor analyses for latent variables at Time 2. Standardized factor loadings are given next to each link from latent variable to indicator variable. Covariances are indicated by curved arrows with *r* values next to them.

indirect effect. Table 3 reports odds ratios and their confidence intervals, as well as significance levels for each of the three indirect effects. Mplus also reports whether indirect effects are significantly different from each other, as well as giving the total effect of the programme on the outcome (all indirect effects, plus the direct effect, added together).

The Reading model

The full model showed that completion of the Year 1 reading programme was significantly associated with differences in grapheme/phoneme knowledge (β = .33, 95% CI [0.09–0.57]), decoding (β = .26, 95% CI [0.09–0.43]), and comprehension (β = .26, 95% CI [0.05–0.47]) at T2. These βs can be interpreted as effect sizes as they represent the difference in standard deviations in the outcome variable between the control and programme groups. The differences in decoding and comprehension, in turn, significantly predicted reading outcome at T3. Differences in G/P knowledge were not associated with reading outcome (Figure 3). Indirect effects from programme to reading via decoding, and from programme to reading via comprehension were significant and significantly greater than the indirect effect via G/P knowledge (Table 3).

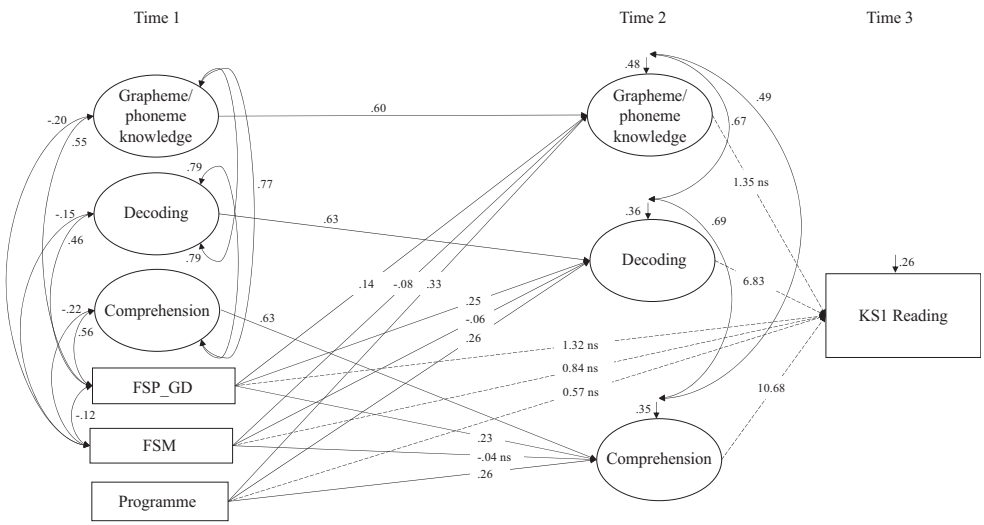

Figure 3. KS1 Reading mediation model. Solid lines denote linear regression analyses; standardized regression weights (betas) are given next to solid lines from continuous predictors (grapheme/phoneme knowledge, decoding, and comprehension), and unstandardized regression weights (effect sizes) are given next to solid lines from binary predictors (FSP_GD, FSM, Programme). Dashed lines denote logistic regression analyses; odds ratios are given next to each dashed line. All links are significant at $p < .05$, unless denoted with ns (not significant). Residuals (1 − multiple squared correlation (r^2)) are given above the short arrows feeding into the endogenous variables. Curved lines indicate covariances. See CFAs (Figures 1 and 2) for factor loadings of the indicator variables for each latent variable (indicated by an oval). FSM = free school meal status, FSP_GLD = Early Years Foundation Stage Profile Good Level of Development.

Table 3. Regression coefficients (unstandardized) and odds ratios for indirect and total effects in the reading and writing models

	Reading model			Writing model		
	Regression coefficient	Odds ratio (OR)	95% CI for OR	Regression coefficient	Odds ratio (OR)	95% CI for OR
Indirect effects						
Programme→G/P knowledge→KS1 outcome	.09	1.09	0.94–1.27	.10	1.10	0.95–1.28
Programme→Decoding→KS1 outcome	.45*	1.57	1.03–2.38	.26*	1.29	1.02–1.65
Programme→Comprehension→KS1 outcome	.32*	1.38	1.06–1.79	.22*	1.25	1.02–1.53
Total indirect effects	.86*	2.31	1.36–4.09	.58*	1.78	1.24–2.55
Direct effect						
Programme→KS1 outcome	–.55	0.57	0.26–1.28	–.04	0.96	0.42–2.17
Total effect (direct + indirect)	.30	1.35	0.68–2.68	.54	1.71	0.82–3.55

Odds ratios are calculated by exponentiating the unstandardized logistic regression coefficients.
*$p < .05$.

The writing model

Replicating the reading model, completion of the Year 1 reading programme was significantly associated with differences in grapheme/phoneme knowledge (β = .34, 95% CI [0.10–0.58]), decoding (β = .27, 95% CI [0.10–0.43]) and comprehension (β = .27, 95% CI [0.06–0.47]) at T2. The differences in decoding and comprehension, in turn, significantly predicted writing outcome at T3. Differences in G/P knowledge were not associated with writing outcome (Figure 4). Indirect effects from programme to writing via decoding and from programme to writing via comprehension were significant and significantly greater than the indirect effect via G/P knowledge (Table 3).

Discussion

Johnson et al. (2019) demonstrated that the Year 1 reading programme had a positive effect on Key Stage 1 literacy tests in England (UK) in a well-executed, large-scale, blind RCT. The next step was to show whether this was the result of a boost in one or more of the skills that the programme was specifically designed to train, rather than an untested 'third' variable, or a Hawthorne effect. Specifically, our aim was to determine whether the effects of the reading programme on Key Stage 1 assessments could be attributed to an increase in grapheme/phoneme knowledge, decoding, and comprehension abilities, as targeted by the programme. The results of the mediation analyses showed that this was the case for decoding and comprehension, but not for grapheme/phoneme knowledge. There were significant indirect effects from the programme to both Key Stage 1 reading and writing via both decoding and comprehension latent constructs. A child in the programme group was more likely to achieve or exceed their expected level at Key Stage 1 reading compared with a control group child by 1.6 times due to an increase in their decoding ability, and by 1.4 times due to an increase in their comprehension abilities. Similarly for writing, children in the programme group were more likely to achieve or exceed their expected level at Key Stage 1 writing compared with the control group by 1.3 times due to an increase in their decoding ability, and by 1.3 times due to an increase in their comprehension abilities. However, the indirect effect from programme to Key Stage 1 reading and writing via grapheme/phoneme knowledge was not significant. Nevertheless, the programme did have a significant effect on grapheme/phoneme knowledge at the end of the programme such that children in the programme group increased their grapheme/phoneme knowledge by 0.33 standard deviations more than those in the control group. The total indirect effect (via grapheme/phoneme knowledge, decoding, and comprehension) from programme to Key Stage 1 reading and writing showed that children in the programme group were 2.31 and 1.78 times, respectively, more likely to achieve or exceed their expected level at Key Stage 1 than the control group children. Such a strong effect is of real educational significance and provides strong support for the integrity of the programme.

It may appear surprising at first glance that Key Stage 1 reading and writing outcomes were not enhanced by the basic skills of grapheme/phoneme knowledge. However, Key Stage 1 reading and writing outcomes focus more on text-level skills than word-level skills such as decoding and spelling. While it is clear that grapheme/phoneme-level knowledge such as PA and LSK is very likely causally related to word-level skills such as decoding and spelling (Hulme et al., 2012), any effects of grapheme/phoneme knowledge on higher level skills such as comprehension and writing are more likely to be indirect, through their effect on word-level skills. The models presented here cannot speak to this possibility

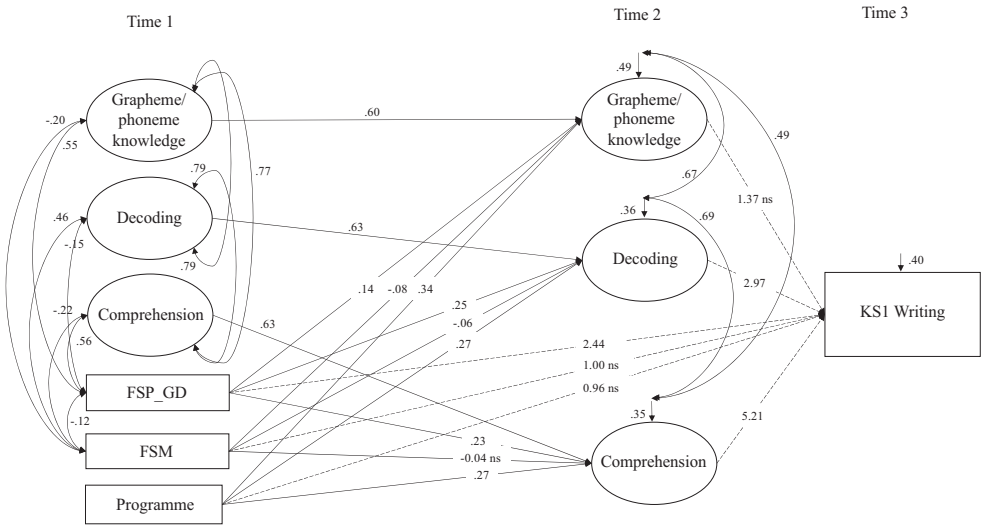

Figure 4. KS1 writing mediation model. Solid lines denote linear regression analyses; standardized regression weights (betas) are given next to solid lines from continuous predictors (Grapheme/phoneme knowledge, decoding, and comprehension), and unstandardized regression weights (effect sizes) are given next to solid lines from binary predictors (FSP_GD, FSM, Programme). Dashed lines denote logistic regression analyses; odds ratios are given next to each line. All links are significant at $p < .05$, unless denoted with ns (not significant). Residuals (1 − multiple squared correlation (r^2)) are given above the short arrows feeding into the endogenous variables. Curved lines indicate covariances. See CFAs (Figures 1 and 2) for factor loadings of the indicator variables for each latent variable (indicated by an oval). FSM = free school meal status, FSP_GLD = Early Years Foundation Stage Profile Good Level of Development.

because there was no testing point between T2 and T3. Additionally, we acknowledge that there is more to grapheme/phoneme knowledge than captured by the grapheme/-phoneme latent variable (measured primarily by letter-sound knowledge). It is possible a more sensitive measure of phonological awareness might have led to a more comprehensive latent variable, which in turn may have explained more variance in the outcomes than the current latent variable.

A main feature of the Year 1 reading programme was the inclusion of both decoding and comprehension activities in the same programme, with the aim of understanding how the programme improves distal educational outcomes. Although early reading instruction necessarily focuses on decoding, national educational outcome tests focus on higher-order skills such as reading comprehension and writing, even in the early years. It is therefore imperative that we understand how programmes aimed at early readers contribute to these more distal educational outcomes. Improving reading comprehension in early readers has typically been approached by oral language intervention, but despite the strong link between oral language and reading comprehension (Hjetland, Brinchmann, Scherer, & Melby-Lervåg, 2017; Lervåg et al., 2018), impact via this route can be hard to demonstrate (Rogde et al., 2019; but see, e.g., Fricke, Bowyer-Crane, Haley, Hulme, & Snowling, 2013; Language and Reading Research Consortium (LARRC) et al., 2019). Reading comprehension interventions, on the other hand, typically target older

readers (Clarke et al., 2010; Davis, 2010; Edmonds et al., 2009; Scammacca et al., 2015) because younger readers do not have sufficient decoding skills to access reading comprehension activities, supporting a developmental model of 'phonics first' and 'comprehension later', as suggested by Suggate (2010). In contrast, the Year 1 reading programme demonstrates that comprehension instruction can be delivered effectively in the early years by integrating it into a programme with decoding instruction. This is consistent with previous work using same ABRA software (e.g., Savage, Abrami, et al., 2009), where decoding and comprehension activities are linked to the same texts, meaning that children learn to decode the material that they are also learning to comprehend. However, previous work has not tested whether these skills acted as mediators to a more distal educational outcome (e.g., a national test). The Year 1 reading programme builds on this work (Savage, Abrami, et al., 2009) by demonstrating that distal educational outcomes are positively affected by the programme through both decoding and comprehension components, suggesting that instruction from both components is effective in raising attainment. The size of the indirect effects of the programme through decoding and comprehension on Key Stage 1 reading was similar (OR of 1.6 and 1.4 respectively) and not significantly different (Wald's chi-square = .37, p = .55), suggesting both are equally important. This was also the case for writing (OR of 1.3 and 1.3 respectively, Wald chi-sq = .06, p = .80).

It is difficult to compare the size of effects found here with previous work because the outcomes in this study are dichotomous and are therefore measured by odds ratios. However, it is notable that children in the programme group were about twice as likely to meet or exceed the expected level of performance in their Key Stage 1 outcomes than the control group. This sizeable effect came from a combination of the uplift in specific skills trained by the programme. Although one can never rule out additional contributions from unforeseen third variables, the results do provide strong evidence in favour of the programme, and against Hawthorne effects.

Demonstrating effects on distal outcomes can be challenging, especially when the distal outcomes contain different contents to the material in the instructed programme (Melby-Lervåg, Hagen, & Lervåg, 2020). The distal measures in this study (Key Stage 1 reading and writing) were based on teacher assessments of pupils' abilities in reading and writing collected up to a year after the programme had finished. It is unlikely therefore that there would have been much overlap between the content of the Year 1 reading programme and the work on which teacher assessments were based. This is especially the case for the writing outcome since writing was not part of the programme, yet significant positive effects were still observed. Additionally, these assessments were made by Year 2 teachers, who for the most part are different teachers to the Year 1 teachers. Thus, both the teacher and the context differ when the assessments are made. The activities within the Year 1 reading programme focused on evidence-based skills and strategies that were practised repeatedly in the context of different texts. This encourages the development of skills and strategies independent of content and favours an explanation based on the application of learned skills and strategies rather than a content-based mechanism. This explanation is consistent with evidence that writing benefits from similar skills as reading comprehension, since writing was not taught in the programme, but some skills that have been linked to both reading comprehension and writing were, for example, comprehension monitoring (Limpo et al., 2014), summarizing (Mason et al., 2006), vocabulary (Savage et al., 2017), and story structure (Arrimada et al., 2019; Pinto et al., 2016; Spencer & Petersen, 2018).

Limitations

The aim of the current study was to understand the impact of the Year 1 reading programme on educational outcomes, which necessitated the use of national educational outcomes over traditional standardized tests. Thus, the outcomes in both models were teacher assessments rather than standardized tests. Although it is known that effect sizes based on non-standardized tests are often larger than those based on standardized tests (Davis, 2010; Edmonds et al., 2009; Scammacca et al., 2015), the Key Stage 1 outcomes reported here are based on a highly structured procedure practiced nationally to assess all pupils. Furthermore, the content of the teacher assessments was skills based and assessed up to a year after the programme had finished. Thus, although the outcomes were not standardized tests, neither were they experimenter-generated tests linked to the content of the programme. Lastly, the teacher assessments were not conducted by the teaching assistants who delivered the programme, they were conducted by a different year group teacher when the children were in the next year up. These are some reasons to be confident that the outcomes are not biased and of significant impact. A more robust test of the significance of the impact would be to see whether the programme still has impact at the next Key Stage in education, when these children sit examinations.

The results of the study do not distinguish between the types of activities that are more or less effective for either decoding or comprehension – both components were delivered via a range of activities. However, this study was the first step in testing whether both components contribute to distal educational outcomes. Future research could take a finer grained approach to determine whether all activities are equally useful in this respect.

Implications for practice

The findings of the current study provide strong evidence that reading comprehension can be successfully delivered to beginning readers who are still learning to decode text. Compared with the taught activities for grapheme/phoneme knowledge, the taught activities for decoding and comprehension focused more heavily on the application of knowledge and practice of learned procedural skills rather than the acquisition of knowledge per se (Supporting Information). It is worth noting that the knowledge-based learning from the grapheme/phoneme component of the Year 1 reading programme is additionally taught elsewhere in school through phonics instruction. However, the results indicated that improvements to Key Stage 1 reading and writing were mediated though decoding and comprehension, not grapheme/phoneme knowledge. This suggests that there is a clear role for opportunities for children to apply their knowledge and practice-learned procedural skills and acquire the phonic knowledge already taught in school and in the programme.

The findings also demonstrated that reading comprehension had a measurable and significant impact on beginning readers, whether or not they have weaknesses in decoding, oral language, or both, since the programme was delivered to children of all abilities, rather than children with language or literacy weaknesses, as is often the case for comprehension research (Bowyer-Crane et al., 2008; Clarke et al., 2010; Edmonds et al., 2009; Scammacca et al., 2015). The implication for practice is that beginning readers can benefit in a meaningful way within an educational context when reading comprehension should be introduced in parallel, and well integrated, with decoding instruction.

Conclusion

The findings of the current study are consistent with a body of research showing that comprehensive reading programmes can be effective in promoting growth in more than one component of reading (Abrami, Savage, Wade, Hipps, & Lopez, 2008; Bailey et al., 2017; Savage, Abrami, et al., 2009; Savage et al., 2010). Our findings add to this body of research by showing that improved performance on distal educational outcomes can be achieved through improvements in the trained components for those children who participated in the Year 1 reading programme. These effects were demonstrated with children of all abilities. Effects on distal measures can be hard to demonstrate in large-scale RCTs, yet the effects demonstrated here are sizeable and are of practical significance in helping children reach the expected level of literacy by the end of Key Stage 1. Furthermore, the significant effects demonstrated here were obtained from a quality-scaled school-based blind RCT with a robust assessment model (Huber–White error model) with latent variables, and are therefore replicable.

The current study joins a modest but growing number of studies that have used mediation analyses to understand which components of programmes and interventions are responsible for positive outcomes (Clarke et al., 2010; Hulme et al., 2012; Language and Reading Research Consortium (LARRC) et al., 2019; Melby-Lervåg et al., 2020). This approach allows intervention and programme research to focus on the components that are responsible for positive effects, and to discontinue components that are ineffective. Given that children at the start of the study already had some ability in decoding and comprehension, the findings cannot claim to be truly causal in this respect. However, the findings presented here alongside the main effects of the well-executed RCT reported in Johnson et al. (2019) are at least consistent with a causal interpretation that Key Stage 1 reading and writing outcomes are (in significant part) caused by the improvements in both comprehension and decoding ability targeted by the Year 1 reading programme.

Acknowledgements

Janet Vousden, Anna Cunningham, and Clare Wood now work at Nottingham Trent University, Helen Johnson now works at Lomond Educational Psychology, Sam Waldron now works at Birmingham University, Sabrina Ammi now works for Deloitte, and Claire Pillinger and Robert Savage now work at University College London. The research reported in this article was based on a registered RCT (https://doi.org/10.1186/ISRCTN18254678) funded by The Education Endowment Foundation (EEF), UK. The EEF evaluation has been reported elsewhere (Johnson et al., 2019; McNally et al., 2016). We thank the Literacy Group at the Centre for Research in Psychology, Behaviour and Achievement for their help with running the trial. We thank Sandra McNally and Jenifer Ruiz-Valenzuela at the Centre for Vocational Education Research, London School of Economics and Heather Rolfe at the National Institute of Economic and Social Research for leading the impact and process evaluation. We thank Anne Wade and Philip Abrami, and others at Concordia University's Centre for the Study of Learning and Performance (CSLP) for their technical support. We thank participating schools and teaching assistants. We thank Morag Stuart for implementing the non-Web-based version of the programme.

Conflicts of interest

All authors declare no conflict of interest.

Author contributions

Claire Pillinger (Data curation; Investigation; Project administration) Sabrina Ammi (Data curation; Investigation; Project administration; Resources; Supervision) Clare Wood (Conceptualization; Funding acquisition; Investigation; Methodology; Project administration; Supervision; Writing – review & editing) Rob Savage (Conceptualization; Funding acquisition; Methodology; Resources; Supervision; Writing – review & editing) Anna J. Cunningham (Formal analysis; Writing – original draft; Writing – review & editing) Janet Vousden, Ph.D. (Conceptualization; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Supervision; Writing – original draft; Writing – review & editing) Sam Waldron (Data curation; Investigation; Project administration; Resources; Supervision) Helen Johnson (Data curation; Investigation; Project administration; Resources; Supervision).

Data availability statement

The data that support the findings of this study are available from the Education Endowment Foundation and the Office for National Statistics. Restrictions apply to the availability of these data, which were used under licence for this study. The data are not publicly available due to privacy or ethical restrictions.

References

Abrami, P. C., Lysenko, L., & Borokhovski, E. (2020). The effects of ABRACADABRA on reading outcomes: An updated meta-analysis and landscape review of applied field research. *Journal of Computer Assisted Learning*, *36*(3), 260–279. https://doi.org/10.1111/jcal.12417

Abrami, P. C., Savage, R., Deleveaux, G., Wade, A., Meyer, E., & Lebel, C. (2010). The design, development, testing and dissemination of evidence-based educational software. In P. Zemliansky, & D. M. Wilcox (Eds.), *Design and implementation of educational games: Theoretical and practical perspectives* (pp. 168–187). Hershey, PA: IGI Global.

Abrami, P. C., Savage, R., Wade, C. A., Hipps, G., & Lopez, M. (2008). Using technology to assist children learning to read and write. In T. Willoughby & E. Wood (Eds.), *Children's learning in a digital world* (pp. 129–171). Oxford, UK: Blackwell Publishing.

Arias, S. A., Dumas, O., Sullivan, A. F., Boudreaux, E. D., Miller, I., & Camargo, C. A. (2016). Substance use as a mediator of the association between demographics, suicide attempt history, and future suicide attempts in emergency department patients. *Crisis*, *37*, 385–391. https://doi.org/10.1027/0227-5910/a000380

Arlinghaus, A., Lombardi, D. A., Willetts, J. L., Folkard, S., & Christiani, D. C. (2012). A structural equation modeling approach to fatigue-related risk factors for occupational injury. *American Journal of Epidemiology*, *176*, 597–607. https://doi.org/10.1093/aje/kws219

Arrimada, M., Torrance, M., & Fidalgo, R. (2019). Effects of teaching planning strategies to first-grade writers. *British Journal of Educational Psychology*, *89*, 670–688. https://doi.org/10.1111/bjep.12251

Bailey, B., Arciuli, J., & Stancliffe, R. J. (2017). Effects of ABRACADABRA Literacy Instruction on children with autism spectrum disorder. *Journal of Educational Psychology*, *109*, 257–268. https://doi.org/10.1037/edu0000138

Berninger, V. W., Vaughan, K., Abbott, R. D., Begay, K., Coleman, K. B., Curtin, G., ... Graham, S. (2002). Teaching spelling and composition alone and together: Implications for the simple view of writing. *Journal of Educational Psychology*, *94*, 291–304. https://doi.org/10.1037/0022-0663.94.2.291

Bowyer-Crane, C., Snowling, M. J., Duff, F. J., Fieldsend, E., Carroll, J. M., Miles, J., … Hulme, C. (2008). Improving early language and literacy skills: Differential effects of an oral language versus a phonology with reading intervention. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *49*, 422–432. https://doi.org/10.1111/j.1469-7610.2007.01849.x

Byrne, B., & Fielding-Barnsley, R. (1989). Phonemic awareness and letter knowledge in the child's acquisition of the alphabetic principle. *Journal of Educational Psychology*, *81*, 313–321. https://doi.org/10.1037/0022-0663.81.3.313

Byrne, B., & Fielding-Barnsley, R. (1991). Evaluation of a program to teach phonemic awareness to young children. *Journal of Educational Psychology*, *83*, 451–455. https://doi.org/10.1037/0022-0663.83.4.451

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology*, *96*(1), 31–42. https://doi.org/10.1037/0022-0663.96.1.31

Castles, A., & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition*, *91*(1), 77–111. https://doi.org/10.1016/S0010-0277(03)00164-1

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest*, *19*(1), 5–51. https://doi.org/10.1177/1529100618772271

Clarke, P. J., Snowling, M. J., Truelove, E., & Hulme, C. (2010). Ameliorating children's reading-comprehension difficulties: A randomized controlled trial. *Psychological Science*, *21*, 1106–1116. https://doi.org/10.1177/0956797610375449

Davis, D. S. (2010). A meta-analysis of comprehension strategy instruction for upper elementary and middle school students. (Publication No. 61405816) [Doctoral dissertation, Vanderbilt University]. ProQuest Dissertations & Theses Global.

Department for Education. (2018). *National curriculum assessments Key stage 1 Teacher assessment frameworks at the end of key stage 1*. 14. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/740343/2018-19_teacher_assessment_frameworks_at_the_end_of_key_stage_1_WEBHO.pdf

Department for Education. (2019). *Phonics screening check and key stage 1 assessments in England, 2019. September, 1–18*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/851296/Phonics_screening_check_and_key_stage_1_assessments_in_England_2019.pdf

Dodd, B., Crosbie, S., McIntosh, B., Teizel, T., & Ozanne, A. (2000). *Preschool and primary inventory of phonological awareness*. London, UK: The Psychological Corporation.

Dunn, D. M., & Dunn, L. M. (2009). *The British Picture Vocabulary Scale* (3rd ed.). London, UK: GL Assessment.

Edmonds, M. S., Vaughn, S., Wexler, J., Reutebuch, C., Cable, A., Tackett, K. K., & Schnakenberg, J. W. (2009). A synthesis of reading interventions and effects on reading comprehension outcomes for older struggling readers. *Review of Educational Research*, *79*(1), 262–300. https://doi.org/10.3102/0034654308325998

Ehri, L. C. (2020). The science of learning to read words: A case for systematic phonics instruction. *Reading Research Quarterly*, *55*(S1), 45–60, https://doi.org/10.1002/rrq.334

Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001). Systematic phonics instruction helps students learn to read: Evidence from the national reading panel's meta-analysis. *Review of Educational Research*, *71*, 393–447. https://doi.org/10.3102/00346543071003393

Elleman, A. M., Lindo, E. J., Morphy, P., & Compton, D. L. (2009). Intervention, evaluation, and policy studies: The impact of vocabulary instruction on passage-level comprehension of school-age children: A meta-analysis. *Journal of Research on Educational Effectiveness*, *2*(1), 1–44. https://doi.org/10.1080/19345740802539200

Forum for Research in Literacy and Language (2012). *Diagnostic test of word reading processes*. London, UK: GL Assessment.

Fricke, S., Bowyer-Crane, C., Haley, A. J., Hulme, C., & Snowling, M. J. (2013). Efficacy of language intervention in the early years. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *54*(3), 280–290. https://doi.org/10.1111/jcpp.12010

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, *7*(1), 6–10. https://doi.org/10.1177/074193258600700104

Hatcher, P. J., Hulme, C., & Snowling, M. J. (2004). Explicit phoneme training combined with phonic reading instruction helps young children at risk of reading failure. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *45*, 338–358. https://doi.org/10.1111/j.1469-7610.2004.00225.x

Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: Guildford Press.

Hjetland, H. N., Brinchmann, E. I., Scherer, R., & Melby-Lervåg, M. (2017). Preschool predictors of later reading comprehension ability: A systematic review. *Campbell Systematic Reviews*, *13*(1), 1–155. https://doi.org/10.4073/csr.2017.14

Hulme, C., Bowyer-Crane, C., Carroll, J. M., Duff, F. J., & Snowling, M. J. (2012). The causal role of phoneme awareness and letter-sound knowledge in learning to read: Combining intervention studies with mediation analyses. *Psychological Science*, *23*(6), 572–577. https://doi.org/10.1177/0956797611435921

Hulme, C., & Snowling, M. J. (2013). Learning to read: What we know and what we need to understand better. *Child Development Perspectives*, *7*(1), 1–5. https://doi.org/10.1111/cdep.12005

Iacobucci, D., Saldanha, N., & Deng, X. (2007). A meditation on mediation: Evidence that structural equations models perform better than regressions. *Journal of Consumer Psychology*, *17*(2), 139–153. https://doi.org/10.1016/S1057-7408(07)70020-7

Johnson H., McNally S., Rolfe H., Ruiz-Valenzuela J., Savage R., Vousden J., Wood C. (2019). Teaching assistants, computers and classroom management. *Labour Economics*, *58*, 21–36. http://dx.doi.org/10.1016/j.labeco.2019.02.006

Juniper Education. (2021). *Juniper Education National Dataset Report, February 2021 – The impact of the Covid-19 pandemic on primary school children's learning (Issue February)*. Retrieved from https://21e8jl3324au2z28ej2uho3t-wpengine.netdna-ssl.com/wp-content/uploads/juniper_folder/Juniper-Education-National-Benchmark-Dataset-Report.pdf

Justice, L. M., Cain, K., Yeomans-Maldonado, G., Lomax, R., O'Connell, A., Pentimonti, J., … Yeomans-Maldonado, G. (2017). Development of comprehension monitoring in beginner readers. *Reading and Writing*, *30*, 2039–2067. https://doi.org/10.1007/s11145-017-9765-x

Kenny, D. A., & Judd, C. M. (2014). Power anomalies in testing mediation. *Psychological Science*, *25*, 334–339. https://doi.org/10.1177/0956797613502676

Kim, Y. S. G., & Schatschneider, C. (2017). Expanding the developmental models of writing: A direct and indirect effects model of developmental writing (DIEW). *Journal of Educational Psychology*, *109*(1), 35–50. https://doi.org/10.1037/edu0000129

Language and Reading Research Consortium (LARRC), Jiang, H., & Logan, J. (2019). Improving reading comprehension in the primary grades: Mediated effects of a language-focused classroom intervention. *Journal of Speech, Language, and Hearing Research*, *62*, 2812–2828. https://doi.org/10.1044/2019_JSLHR-L-19-0015

Larsen, L., Kohnen, S., Nickels, L., & McArthur, G. (2015). The Letter-Sound Test (LeST): A reliable and valid comprehensive measure of grapheme–phoneme knowledge. *Australian Journal of Learning Difficulties*, *20*(2), 129–142, https://doi.org/10.1080/19404158.2015.1037323

Lervåg, A., Hulme, C., & Melby-Lervåg, M. (2018). Unpicking the developmental relationship between oral language skills and reading comprehension: It's simple, but complex. *Child Development*, *89*, 1821–1838. https://doi.org/10.1111/cdev.12861

Limpo, T., Alves, R. A., & Fidalgo, R. (2014). Children's high-level writing skills: Development of planning and revising and their contribution to writing quality. *British Journal of Educational Psychology*, *84*(2), 177–193. https://doi.org/10.1111/bjep.12020

Lonigan, C. J., Burgess, S. R., & Schatschneider, C. (2018). Examining the simple view of reading with elementary school children: Still simple after all these years. *Remedial and Special Education*, *39*, 260–273. https://doi.org/10.1177/0741932518764833

Mason, L. H., Snyder, K. H., Sukhram, D. P., & Kedem, Y. (2006). TWA + Plans strategies for expository reading and writing: Effects for nine fourth-grade students. *Exceptional Children*, *73* (1), 69–89. https://doi.org/10.1177/001440290607300104

McCarty, C., & Ruttle, K. (2010). *Progress in reading assessment*. London, UK: Hodder Education.

McKinnon, D. P. (2008). *An introduction to statistical mediation analysis*. London, UK: Routledge.

McNally, S., Ruiz-valenzuela, J., & Rolfe, H. (2016). ABRA : Online Reading Support, Evaluation report and executive summary. Education Endowment Foundation. https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/ABRA_with_addendum.pdf

Melby-Lervåg, M., Hagen, Å. M., & Lervåg, A. (2020). Disentangling the far transfer of language comprehension gains using latent mediation models. *Developmental Science*, *23*, 1–14. https://doi.org/10.1111/desc.12929

Melby-Lervåg, M., Lyster, S. A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin*, *138*, 322–352. https://doi.org/10.1037/a0026744

Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology*, *40*, 665–681. https://doi.org/10.1037/0012-1649.40.5.665

Muthen, L. K., & Muthen, B. O. (2017). Mplus Version 8 User's Guide. Los Angeles, CA: *Muthen & Muthen*.

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. Reports of the subgroups (NIH Publication No. 00-4754)*. Retrieved from https://www.nichd.nih.gov/sites/default/files/publications/pubs/nrp/Documents/report.pdf

Oakhill, J., & Cain, K. (2012). The precursors of reading ability in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading*, *16*, 91–121. https://doi.org/10.1080/10888438.2010.529219

Oakhill, J. V., Cain, K., & Bryant, P. E. (2003). The dissociation of word reading and text comprehension: Evidence from component skills. *Language and Cognitive Processes*, *18*, 443–468. https://doi.org/10.1080/01690960344000008

Piasta, S. B., & Wagner, R. K. (2010). Developing early literacy skills: A meta-analysis of alphabet learning and instruction. *Reading Research Quarterly*, *45*(1), 8–38. https://doi.org/10.1598/rrq.45.1.2

Pinto, G., Tarchi, C., & Bigozzi, L. (2016). Development in narrative competences from oral to written stories in five- to seven-year-old children. *Early Childhood Research Quarterly*, *36*, 1–10. https://doi.org/10.1016/j.ecresq.2015.12.001

Rogde, K., Hagen, Å. M., Melby-Lervåg, M., & Lervåg, A. (2019). The effect of linguistic comprehension instruction on generalized language and reading comprehension skills: A systematic review. *Campbell Systematic Reviews*, *15*, 1–37. https://doi.org/10.1002/cl2.1059

Rose, J. (2006). Independent review of the teaching of early reading final report. *Independent Review*. Retrieved from www.standards.dcsf.gov.uk/phonics/report.pdf

Savage, R., Abrami, P., Hipps, G., & Deault, L. (2009). A Randomized Controlled Trial Study of the ABRACADABRA Reading Intervention Program in Grade 1. *Journal of Educational Psychology*, *101*, 590–604. https://doi.org/10.1037/a0014700

Savage, R., Abrami, P. C., Piquette, N., Wood, E., Deleveaux, G., Sanghera-Sidhu, S., & Burgos, G. (2013). A (Pan-Canadian) cluster randomized control effectiveness trial of the ABRACADABRA web-based literacy program. *Journal of Educational Psychology*, *105*, 310–328. https://doi.org/10.1037/a0031025

Savage, R., & Carless, S. (2004). Predicting curriculum and test performance at age 7 years from pupil background, baseline skills and phonological awareness at age 5. *British Journal of Educational Psychology*, *74*(2), 155–171. https://doi.org/10.1348/000709904773839815

Savage, R., & Carless, S. (2005). Phoneme manipulation not onset-rime manipulation ability is a unique predictor of early reading. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *46*, 1297–1308. https://doi.org/10.1111/j.1469-7610.2005.01436.x

Savage, R., & Carless, S. (2008). The impact of early reading interventions delivered by classroom assistants on attainment at the end of year 2. *British Educational Research Journal*, *34*, 363–385. https://doi.org/10.1080/01411920701609315

Savage, R., Carless, S., & Erten, O. (2009). The longer-term effects of reading interventions delivered by experienced teaching assistants. *Support for Learning*, *24*(2), 95–100. https://doi.org/10.1111/j.1467-9604.2009.01405.x

Savage, R., Erten, O., Abrami, P., Hipps, G., Comaskey, E., & van Lierop, D. (2010). ABRACADABRA in the hands of teachers: The effectiveness of a web-based literacy intervention in grade 1 language arts programs. *Computers and Education*, *55*, 911–922. https://doi.org/10.1016/j.compedu.2010.04.002

Savage, R., Georgiou, G., Parrila, R., Maiorino, K., Dunn, K., & Burgos, G. (2020). The effects of teaching complex grapheme-phoneme correspondences: Evidence from a dual site cluster trial with at-risk grade 2 students. *Scientific Studies of Reading*, *24*(4), 321–337. https://doi.org/10.1080/10888438.2019.1669607

Savage, R., Kozakewich, M., Genesee, F., Erdos, C., & Haigh, C. (2017). Predicting writing development in dual language instructional contexts: Exploring cross-linguistic relationships. *Developmental Science*, *20*(1), 1–13. https://doi.org/10.1111/desc.12406

Scammacca, N. K., Roberts, R., Vaughn, S., & Stuebing, K. K. (2015). A meta-analysis of interventions for struggling readers in Grades 4–12: 1980–2011. *Journal of Learning Disabilities*, *48*(4), 369–390. https://doi.org/10.1177/0022219413504995

Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C. D., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology*, *96*, 265–282. https://doi.org/10.1037/0022-0663.96.2.265

Schleicher, A. (2019). *PISA 2018: Insights and interpretations*. OECD Publishing. Retrieved from https://www.oecd.org/pisa/PISA2018InsightsandInterpretationsFINALPDF.pdf

Snowling, M. J., Hulme, C., Bailey, A. M., Stothard, S. E., & Lindsay, G. (2011). *Better communication research project: Language and literacy attainment of pupils during early years and through KS2: Does teacher assessment at five provide a valid measure of children's current and future educational attainments?* (Issue November). Retrieved from http://wrap.warwick.ac.uk/45721/

Spencer, T. D., & Petersen, D. B. (2018). Bridging oral and written language: An oral narrative language intervention study with writing outcomes. *Language, Speech, and Hearing Services in Schools*, *49*, 569–581. https://doi.org/10.1044/2018_LSHSS-17-0030

Spybrook, J., Bloom, H., Congdon, R., Hill, C., Martinez, A., & Raudenbush, S. (2011). *Optimal design plus empirical evidence: Documentation for the "optimal design" software*. NewYork, NY: WIlliam T Grant Foundation. http://wtgrantfoundation.org/library/uploads/2015/11/OD-Documentation-V3.pdf

Standards and Testing Agency. (2013). *Early Years Foundation Stage Profile Handbook 2014*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/301256/2014_EYFS_handbook.pdf

Standards and Testing Agency (2016). *Interim teacher assessment frameworks at the end of key stage 1 Key stage 1* (Issue December 2016). Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/461547/Interim_teacher_assessment_frameworks_at_the_end_of_key_stage_1_PDFA.pdf

Stuart, M. (1999). Getting ready for reading: Early phoneme awareness and phonics teaching improves reading and spelling in inner-city second language learners. *British Journal of Educational Psychology*, *69*, 587–605. https://doi.org/10.1348/000709999157914

Suggate, S. P. (2010). Why what we teach depends on when: Grade and reading intervention modality moderate effect size. *Developmental Psychology*, *46*, 1556–1579. https://doi.org/10.1037/a0020612

Torgerson, C., Brooks, G., Gascoine, L., & Higgins, S. (2019). Phonics: Reading policy and the evidence of effectiveness from a systematic 'tertiary' review. *Research Papers in Education*, *34*, 208–238. https://doi.org/10.1080/02671522.2017.1420816

Whitehurst, G. J., & Lonigan, C. J. (1998). Child Development and Emergent Literacy. *Child Development*, *69*, 848–872. https://doi.org/10.1111/j.1467-8624.1998.tb06247.x

Williams, J. P., Brooke, S. K., Lauer, K. D., Hall, K. M., & Pollini, S. (2009). Embedding reading comprehension training in content-area instruction. *Journal of Educational Psychology*, *101* (1), 1–20. https://doi.org/10.1037/a0013152

Williams, J. P., Pollini, S., Nubla-Kung, A. M., Snyder, A. E., Garcia, A., Ordynans, J. G., & Atkins, J. G. (2014). An intervention to improve comprehension of cause/effect through expository text structure instruction. *Journal of Educational Psychology*, *106*(1), 1–17. https://doi.org/10.1037/a0033215

Yeung, S. S.-S., & Savage, R. (2020). Teaching grapheme-phoneme correspondences using a direct mapping approach for at-risk second language learners: A randomized controlled trial. *Journal of Learning Disabilities*, *53*(2), 131–144. https://doi.org/10.1177/0022219419894563

Zhao, X., Lynch, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research*, *37*, 197–206. https://doi.org/10.1086/651257

Received 9 March 2021; revised version received 11 June 2021

Supporting Information

The following supporting information may be found in the online edition of the article:

Supplementary Material. Further detailed information about the study design (School characteristics, the Year 1 Reading Programme contents, Teaching Assistant characteristics, Training, Treatment Integrity, and the English National Assessment Key Stage 1 outcomes) and Missing data analysis.