Evaluation of Bug Club: a randomised control trial of a whole school primary aged reading programme

Carroll, Catherine, Hurry, Jane, Grima, Grace, Hooper, Alistair, Dunn, Krystina and Ahtaridou, Elpida

Abstract

Literacy levels for children leaving primary school, internationally, remain a cause for concern. There is a limited evidence base for whole school reading programmes and resources in the first two years of primary school compared to studies of interventions for children with reading difficulties. This study used a longitudinal randomised control trial (RCT) design to evaluate the efficacy of Bug Club; a whole school reading programme. Data from1273 pupils in Years 1 and 2 from 30 schools in the United Kingdom were analysed. A new cohort of 1425 Year 1 pupils was recruited in the second year. Over two terms, the experimental group made 1.75 more standardized points progress in reading compared to the control group, a small but significant effect. A replication study in the second year did not elicit the same significant gains. Over five terms, experimental pupils were 11 months ahead on age equivalent reading scores, relative to their chronological age. This study is the first of its type to evaluate the efficacy of a whole school reading programme in the United Kingdom. The findings contribute to an emerging understanding of what pupil progress might be expected as a result of a whole school reading programme.

Key words: reading programme, literacy, primary school, randomised control trial,

Introduction

Literacy attainment for children in England

Literacy levels for children and young adults in England continue to remain a cause for concern. Interim 2018 national attainment figures in England at the end of primary school (age 11) show that 75% of children achieved the desired standard in reading (DfE, 2018). Therefore, a quarter of eleven year olds in England left primary school without the literacy skills deemed necessary to effectively access the secondary school curriculum. A recent report from the Organisation for Economic Cooperation and Development (OECD) showed how this disparity continues into the 16-24 age group with almost 30% failing to attain a level 2 (expected standard) qualification in literacy in England (Kuczera, Field, and Windisch 2016). Moreover, the same OECD report found that unlike most developed OECD countries, young adults in England perform no better than older adults in the 55-65 age group.

The long term impact of low literacy levels for individuals and society more broadly are well documented. Unemployed adults are twice as likely to have a weak literacy skills as those in employment (Desjardins et al. 2013). Adults with lower levels of literacy are more likely to experience poorer physical health (Pigmone and DeWalt 2006). The costs to the public purse of failing to read at primary school to the age of 37 have been estimated at between £44,797 and £ 53,098 for each individual (Gross 2006). This data clearly shows the importance of improving literacy skills at all ages, but especially as young as possible to prevent the cumulative impact for an individual of a failure to achieve proficiency and confidence in reading in the early years. Whilst there is a wealth of research on some instructional practices,

such as the importance of explicit phonics teaching, there is little research evidence available for schools to make evidence informed choices about the selection of whole school reading programmes, which include such things as reading resources.

Bug Club Reading Programme

Bug Club is a whole school reading programme, based on the principles of guided reading and synthetic phonics. It is a commercial programme designed to be attractive to schools because it addresses key dimensions of the school's literacy curriculum relating to appropriate reading material and phonics resources, and their use. It is offered as part of, rather than in addition to, standard literacy provision.

The pedagogical principles and activities included in the Bug Club programme reflect many of the evidence informed approaches that have been shown to improve reading in the early years of school. These approaches include: the development of pupils' speaking and listening skills; a balanced and engaging approach to developing reading, which combines both decoding and comprehension skills; a systematic approach to phonics teaching and supporting pupils to develop their own comprehension monitoring skills (Higgins et al., 2017). Moreover, by focusing on greater collaboration between home and school, the programme draws on research which highlights that parent involvement has a positive effect on children's reading acquisition (Scanlon, Anderson and Sweeney 2016, Sénéchal, and Young 2008).

The challenges of researching and assessing the effectiveness of reading programmes in the early years of primary school

Three obstacles, in particular, present challenges in researching and assessing the effectiveness of reading programmes in the early years of primary school. The first relates to the well documented challenges of undertaking RCTs in schools more generally (Hanley et al. 2016). These challenges include, for example, to what extent the classroom environment with all its multiple interactions can allow for the 'standardisation' required of interventions in RCT

approaches. Notwithstanding the issues of fidelity to the intervention, the short-term nature of RCTs, at best lasting for one year in terms of reading, do not always allow for sufficient time to embed effective changes in teacher practice in order to allow the 'best circumstances' for the intervention to succeed.

Secondly, although there exists a large number of RCTs and meta-analyses studies on reading programmes internationally, these are predominantly targeted at small groups and individual pupils who have or are at risk if falling behind in literacy. Thirdly, and most pertinent to this study is that, the whole school reading programmes that have been evaluated, differ to a lesser or greater extent in their content, structure, pedagogy and research design which can make comparing studies problematic. Slavin and colleagues addressed this challenge in their review of effective reading programmes for pupils in the first two years of primary school by categorising the studies into four approaches to improving reading with this age group (Slavin et al. 2009). The four approaches were: reading curricula, instructional technology, instructional process programmes and a combination of curricular and instructional process. Their review of 63 beginning reading studies (randomised or matched control with a minimum duration of twelve weeks) found that instructional process programmes, with a focus on alterations to daily teaching practice, resulted in children making the most progress in reading. In addition, cooperative learning and phonics instruction were a common element of more effective programmes.

Using Slavin et al.'s (2009) categorisation of reading approaches, the Bug Club programme is closest to the curricula approach with its emphasis on the provision of learning and teaching resources and an element of professional development in the form of a teacher toolkit. Table 1 provides a summary of the two specific studies; the Open Court Reading and The Reading Street programmes, included in the Slavin review, that have evaluated reading programmes that might also be defined as using a curricular approach and therefore share sufficient characteristics with Bug Club for comparison (Borman, Dowling and Schneck 2007, Wilkerson, Shannon and Herman 2007). Both programmes focus on decoding and comprehension across the primary years, but in addition, Open Court Reading includes a greater emphasis on writing.

Teachers, for both programmes, were provided with guidance, reading and learning resources for pupils in the classroom and at home. Despite the different measures used to assess vocabulary, decoding and comprehension, there were some trends with findings across both of the studies relevant to age range in this study. Both studies reported similar effect sizes for reading comprehension (0.17 and 0.15 respectively) and neither study found any significant differences between the intervention and controls groups for decoding. Both studies were undertaken for one year only. . More recently, a multiyear (K – 5) scale up effectiveness trial of the Open Court Reading programme with 49 elementary schools (4,500 pupils and over 1000 teachers) in seven districts across the United States (US) found no statistically significant main effects on all pupils' reading performance in the first year of the study and a small negative effect (d = -.09) in the second year (Vaden-Kiernan et al. 2018).

Table 1 here

The present study

There are currently very few RCTs that have evaluated the impact of whole school reading programmes for children in the first years of primary school. Moreover, the findings from these few studies are inconclusive as to the efficacy of this approach. There is some evidence of a positive impact for reading comprehension but not for decoding. Reaching tentative conclusions are complicated by the variations in programme content emphasis, the different measures used and the lack of longitudinal research. This study evaluated the efficacy of the Bug Club whole school reading programme on the reading, spelling and vocabulary skills of pupils in the first two years of primary school in England and Northern Ireland. Specifically, the study sought to investigate if pupils following the Bug Club programme made more progress in literacy compared to a control group and if pupil characteristics influenced the impact of Bug Club. A longitudinal (five terms) and randomised control trial (RCT) design was used to evaluate the programme.

Methods

Research design and sample

A randomised clustered control trial was implemented. Power analyses was run in Stata, Clustersamps with an alpha level = .05, power of 80%, assuming an effect size (d) of 0.224 and

an intraclass correlation (ICC) of 0.0225, estimating a required sample size of 22 schools (11 intervention, 11 control). A final sample of 30 schools from England and Northern Ireland was achieved from an initial recruitment of 42 schools expressing interest in the Bug Club reading programme. The 42 schools were matched in pairs on the basis of: percentage Level 2 at KS1; geographical location (urban and rural); class size, percentage of Free School Meals (FSM – a measure of poverty, see below) and English as an Additional Language (EAL). The 21 pairs were randomly allocated to either Bug Club or control schools but due to poor matching on the overall sample on EAL two pairs were subsequently swapped to improve the overall match between intervention and control schools. Following the drop out of six control schools and the consequent loss of their paired Bug Club school, the final sample comprised 15 Bug Club schools and 15 control schools. In the second project year the control schools adopted Bug Club.

Pupils were assessed using the InCAS reading assessment battery (see Measures). In January 2015, Year 1 (Y1) and Year 2 (Y2) pupils were baseline tested in the original Bug Club schools (groups A and B in Table 2) and waiting list control schools (groups C and D). Bug Club schools then implemented Bug Club. Pupils were reassessed in July 2015 (A1). In the second year of the study (January 2016, A2) the control schools adopted Bug Club, the Y1 children from the previous year, now in Y2 were reassessed (n=656) and 769 new pupils were recruited and assessed in the original Bug Club (group E + a small number of new Y2) and control schools (group F + a small number of new Y2). Pupils in Y1 or Y2 at A2 were reassessed in July 2016 (A3).

Table 2: here

This enabled the following four sets of analyses:

1) Baseline to A1, comparing Bug Club (groups A and B) with controls (groups C and D) (RCT)

2) Baseline to A2, comparing Bug Club and control pupils who were Y1 at baseline (group A with group C) (RCT)

3) Replication of 1 using a waiting list design, comparing control groups in 2015 (C and D from baseline to A1) with pupils in the same schools in 2016, now using Bug Club, (E and F from A2 to A3)

4) Follow up of Bug Club pupils over five terms, baseline to A3 (group A).

The numbers of participating pupils for whom Reading Standardised was available at Baseline (recruited Jan. 2015) or A2 (recruited Jan. 2016) are summarised in Table 3.

Table 3 here

At A1 overall attrition was 11.7%, 7.4% in the control group and 15.9% in the Bug Club group which was largely accounted for by an agreed reduction in one Bug Club school which had oversampled. Attrition between baseline and A2 was only relevant to the Year 1 children recruited in 2015 and was 7.5% in Bug Club schools and 12.2% in control schools. From A2 to A3, this Bug Club group dropped from 320 to 279 (12.8%). Overall attrition from A2 to A3 was 10.9%, 8.8% in control schools and 12.8% in Bug Club schools. There was no statistically significant relationship between attrition at any of these points and either the relevant baseline measure of Reading Standardised or demographic variables uptake of free school meals (an indicator of poverty, see below). A meta-analysis of 35 RCT studies in education reported an average attrition rate of 16% with variation in average attrition rates in the experiment (14%) and control groups (19%) (Valentine and McHugh 2007). Thus the attrition rates in this study reflect those of other RCTs in education.

Bug Club and control groups were compared at baseline and A2 with respect to pupil characteristics (gender, year group, English as an additional Language (EAL), Pupil Premium (PP) and were found to be well matched. The only significant difference was for PP at baseline, (chi-square $(1, N = 1415) = 5.71, p <.017$) where the control group had 20.1% PP compared to 15.3% in the Bug Club group. At the start of this study children living in low income families (roughly 15% of school children nationally) were eligible for free school meals and this has been a widely used proxy measure for poverty in England. During the study, free school meals were superseded by the PP, which essentially applied the same eligibility criteria.

Measures

The InCAS assessment for 5-11 year olds, developed by the Centre for Evaluation and Monitoring (CEM) at the University of Durham was used to test the pupils' reading, spelling and developed ability at baseline, A1, A2 and A3. The InCAS programme is a widely used computerised assessment in English schools, particularly efficient for large scale testing, with age standardised norms and satisfactory reliability and validity (Merrell and Tymms 2007).

Rasch measurement (Bond and Fox 2001) was used to create equal interval scales for each of the InCAS scales and subscales. It effectively provides blind assessment. At baseline children were assessed in small groups, with a supervising adult (usually their classroom teacher or assistant), as advised by the CEM team. InCAS provides an assessment of overall reading ability (Reading Standardised) comprised of subscales for word recognition (select the target word from a choice of five), word decoding (select the target non-word from a choice of five), reading comprehension (selection of word from a choice of three that fits within a passage of text), spelling (single word spelling) and Developed Ability (subscales for picture vocabulary (select word given orally from a set of five pictures) and non-verbal ability (pattern recognition

in a set of dots). Subscales are reported as age equivalents in years, raw and standardised scores are not available. Age equivalent scores were derived from a large national sample (Merrell and Tymms 2007). The InCAS data provides two standardised assessment scores, Reading and Developed Ability, with a mean of 100 and standard deviation of 15. Data on age and gender are collected as part of the InCAS package.

School level data were collected on Free School Meals (FSM) and EAL.pupil level data were collected from schools on Year group, PP andEAL.

Intervention

Bug Club is a whole school reading programme, based on the principles of guided reading and synthetic phonics. It is offered as part of, rather than in addition to, standard literacy provision. It is used primarily in Key Stage 1 in over 5000 schools in the United Kingdom (UK) and internationally in Australia and the Middle East. It combines fiction and nonfiction books with both print and online reading which can be supported by parents and carers in the home. It is designed to help children learn to read by providing carefully levelled books that "children would choose themselves". The Bug Club and Phonics Bug materials are designed to be used as follows:

1. Every book is accompanied by a comprehensive set of teacher notes, which enable the teacher to run a guided reading session. Typically, six children who are reading at the same level will join the teacher in a reading and discussion of the text, with their discussion guided by the notes provided. This usually lasts for 20–30 minutes and enables children to unpick the text and develop their comprehension. Photocopiable activities are also provided for each book, which the teacher can use if they wish as a follow-up to the guided reading session. During the guided session each child will have a copy of the text either in print or digitally via a tablet or laptop. Within Bug Club and Phonics Bug a teacher can assign a book digitally to a child's account for them to read

at home, or provide the child with a print book to read at home which is at the correct level. If teachers assign the digital version of the text, then comprehension/phonics questions are delivered to the child through the text, which help to build understanding while reading independently. The results of these questions are reported to the teacher for formative assessment purposes. Both print and digital copies include notes for parents/carers to support their children in independent reading.

2. Phonics Bug for whole school teaching contains interactive whiteboard lessons to be delivered daily throughout reception and year 1 (4–6 years old). The lessons follow a structured progression of phonic sounds to build children's reading fluency. Each lesson examines the sound, reading, spelling, writing and follow-up activities supported by audio, and in some cases video, to demonstrate and deliver the lesson. Each lesson is supported by a suite of games. These can be played by the class on the interactive whiteboard to help embed and formatively assess knowledge of phonic sounds.

Teachers received training in the implementation of Bug Club, described in the introduction, shortly before implementation. Bug Club is designed for use as part of schools' standard literacy provision, and to enhance the quality of provision, rather than as additional teaching time. To document level of use of Bug Club, teachers in Bug Club schools were asked to report their level of implementation of Bug Club for one week, monthly, in the original Bug Club schools from February to June 2015 and in the novice Bug Club schools (formerly the control schools) February to June 2016. They were asked about the materials they used in that week, how often they used them and in what contexts, using the scoring of 4=every day; 3=2-3 times per week; 2=once a week; 1=not at all. Usage was classified under six groupings, as follows (scores in

each group were weighted by number of variables, thus the range for each group was maximum = 4, minimum = 1):

1. Use of print resources: Bug Club books and phonics books in class and out of class, use of teacher support materials (5 variables).

2. Use of online resources: Bug Club books and phonics books in class and out of class, use of teacher support materials (5 variables).

3. Contexts where Bug Club print resources used: Guided reading, independent reading, home reading, materials to teach (4 variables).

4. Contexts where Bug Club online resources used: Guided reading, independent reading, home reading, materials to teach (4 variables).

5. Contexts where Bug Club phonics resources used: Guided reading, independent reading, home reading, materials to teach (4 variables).

6. Use of non Bug Club resources: books for Guided reading print, online and phonics, books for independent reading, home reading, teacher support materials for print, phonics and spelling and grammar (8 variables).

From baseline to A1 62 of the 69 teachers in Bug Club schools completed all five diaries and results are displayed in Table 4. There was substantial variation in the extent to which teachers used Bug Club resources, and in which contexts. For example, two teachers reported not using Bug Club print resources at all and using Bug Club online resources very little, whilst two teachers reported using print resources daily in a range of contexts (inside and outside class) and using online resources, including Phonics Bug, two to three times weekly (inside and outside class). This variation was observed between schools but also between teachers in the same school. The average score on the use of the range of Bug Club print resources over the five months was just over once a week. Diary data was also examined for the period A2 to A3. Unfortunately during this period a considerably smaller number of teachers completed the

diaries and they did so less consistently. Only 36 teachers in the 15 control schools that started using Bug Club at A2 completed at least one month of the diary by A3 and only three teachers completed all five months. This data was insufficiently complete to provide a reliable estimate Bug Club usage.

Table 4: Here

Control

In common with the Bug Club schools, control schools in England delivered the Key Stage 1 English Programmes of Study introduced in September 2014. The programmes comprised of: spoken language; reading (word reading and comprehension) and writing (transcription, composition and vocabulary, grammar and punctuation (DfE 2013). The control school in Northern Ireland delivered the Language and Literacy *area of learning*; one of six areas taught as part of the primary curriculum (CCEA 2007).

Ethics

Initial interest in the project was indicated by the schools themselves, following full briefing on Bug Club and the study, with the understanding that they would be randomly allocate to receive Bug Club in the first or second year of the study. Classroom teachers were approached individually to give informed consent. Parents were given information on the project and given the opportunity to withdraw their children from the study. Every member of the research team going into schools to collect data had a current DBS clearance provided to the school prior to the visit. Data on both children and teachers was anonymised for storage and reporting, maintaining confidentiality. The content of Bug Club is consistent with children's normal reading curriculum and is based on published principles of good practice. The InCAS assessments are frequently used by UK schools to inform them of individual and group progress in key curriculum area.

Analysis strategy

Differences between Bug Club and control groups were tested for statistical significance using multi-level hierarchical regression in Stata 15, with child and school levels to account for baseline school level differences, and controlling for baseline variables to account for individual differences, as follows: reading/spelling measure, auto-regressing the relevant baseline score on the relevant outcome; PP, EAL, gender and Year group. Parsimonious models were selected to minimise levels of missing data and to make interpretation of the results simpler. All child level variables were initially entered into the regressions in the relevant models but were excluded from the final model if they failed to reach a probability at any point of the analysis at or below $p = .10$. Cohen's d effect sizes were calculated by dividing gain scores on the various reading and spelling measures by their pooled standard deviation.

Results

Bug Club and control groups were similar at baseline on the standard scores for Reading and Developed Ability and samples were close to the national average standard scores (mean 100) for both scales (Table 5). The average age of children at baseline was 6.45 years for both groups

Table 5: Here

Impact of Bug Club by condition

The impact of Bug Club from baseline to A1, baseline to A2 and A2 to A3 was estimated for the standardised reading measure, its subscales, spelling and picture vocabulary (Table 6). Before accounting for school level variance pupils in Bug Club schools made significantly more progress on the reading standardised measure and all of its subscales. However school level variance (intraclass correlation) accounted for 13% of the variance on reading standardised, making it advisable to control for this in a multilevel model. In this model, from baseline to A1,

pupils in the Bug Club schools made significantly more progress on the standardised reading measure than pupils in control schools (1.75 standardised points, es = .18) and significantly more progress in the subscales: word decoding, comprehension and spelling. At A2, for those pupils experiencing Bug Club for three terms (Y1 at baseline moving to Y2 at A2), whilst pupils in Bug Club schools made significantly more progress on Reading Standardised before accounting for school level variation (es .16), in a multilevel model the advantage of the Bug Club group on Reading Standardised (1.92 standardised points, es .16) failed to reach statistical significance but was significant in the word decoding subscale (six months more progress over 12 months. A significant impact of Bug Club on reading progress was not found in the replication study, comparing Y1 and Y2 control children from baseline to A1 with Y1 and Y2 children in the same schools over the same time period in the following year, when their schools had adopted Bug Club (A2 to A3). However, this Bug Club group did make significantly more progress in picture vocabulary than their same school controls from the previous year, with Bug Club children making an average two months more progress over five and a half months than control children.

Table 6: Here

Pupil and school characteristics and progress in literacy

To investigate whether or not Bug Club was more effective for some groups of children than others in terms of reading progress, multilevel regression analyses were conducted to look for any interaction effects. Interaction effects between pupil characteristics and Bug Club were tested from baseline to A1, baseline to A2 and A2 to A3. Interaction effects were explored at the pupil level between Bug Club and gender, Year group, PP and EAL. No pupil level interaction effects were found once school level variation was accounted for, for gender, Year

group or EAL, in other words, these pupil level characteristics made no difference to the effectiveness of Bug Club.

However, a significant interaction effect was found for school level FSM at A1 (Table 7). Bug Club had a greater impact (relative to control children) on the reading gains (Reading Standardised) of children in schools with a larger percentage of children on FSM (Table 7). The interaction effect was not apparent at A1 for pupil level PP, although, comparing subgroups on Reading Standardised, for children on FSM Bug Club children scored significantly higher than control children (Bug Club $M = 104$, $SD = 17$; Control $M = 97$, $SD = 14$; $t(234)$, -3.66, p<.001, $d = 0.48$), for children not on FSM Bug Club children did not score significantly higher than control children (Bug Club $M = 110$, $SD = 15$; Control $M = 109$, $SD = 15$; $t(1,020)$, -0.80, p=.164, $d = 0.07$).

Table 7: Here

At A2, a significant interaction effect was found for pupil level PP on Reading Standardised for those pupils experiencing Bug Club for three terms (Y1 at baseline moving to Y2 at A2) (Table 8). Comparing subgroups on Reading Standardised at A2, for children on FSM Bug Club children scored significantly higher than control children (Bug Club $M = 105$, $SD = 16$; Control $M = 96$, $SD = 16$; $t(112)$, -8.79, p=.004, $d = 0.55$), for children not on FSM Bug Club children did not score significantly higher than control children (Bug Club $M = 110$, $SD = 15$; Control $M = 110$, $SD = 15$; $t(468)$, -0.194, p=.846, $d = 0.02$). At A2 the interaction effect was not apparent for school level FSM.

Table 8: Here

For the replication study, neither the interaction effect for pupil level PP nor school level FSM and condition for the outcome Reading Standardised, reached statistical significance (Table 9). Comparing subgroups on Reading Standardised at A3, for children on FSM Bug Club children

did not score significantly higher than control children (Bug Club $M = 98$, $SD = 17$; Control $M = 97$, $SD = 14$; t(234), -0.965, p=.347, $d = 0.13$), for children not on FSM Bug Club children did not score significantly higher than control children (Bug Club $M = 109$, $SD = 15$; Control $M = 109$, $SD = 14$; t(995), .442, p=.958, $d = -0.03$).

Table 9: Here

The impact of Bug Club over five terms

One group of Bug Club pupils (Group A) could be followed up over five terms to investigate the impact of Bug Club on literacy over time. This group were in Year 1 in 2014-15 and aged 5 years, 11 months when first assessed in January 2015. Their reading age equivalent at that time was one month below the norm for the test (5y 10m) (Table 10). Following the introduction of Bug Club to their classrooms they overtook reading age norms by A1 and although progress slowed subsequently, after five terms (1y 6m) of Bug Club they were 11 months ahead of the norm Reading Age. Children on PP made slightly greater progress in their reading (2y 8m) over the five terms than children not on PP (2y 7m), but this did not approach statistical significance (t(282)= -.508, p=.613). Boys made slightly greater progress in their reading (2y 8m) over the five terms than girls (2y 6m), but this did not reach statistical significance (t(283)= 1.422, p=.156).

Table 10: Here

Bug Club pupils made a spurt of progress from baseline to A1, slowing from A2 to A3. Figure 1 shows the mean for gains in each assessment during the two periods for those pupils followed over five terms. From baseline to A1, in all but the spelling test, the pupils were making over a year's progress in two terms. This was also demonstrated by the statistically significant effects that were found at A1 and A2 between the Bug Club and control groups on many of the

assessments. From A2 to A3 the pupils were continuing to make expected but not accelerated progress.

Figure 1: Here

Discussion

This last section summarises the main findings of the evaluation, explores possible explanations of the findings and implications for the reading curriculum in the early years of primary school. It ends with some reflections on the RCT approach in schools and education research more broadly based on our learning from the strengths and limitations of the Bug Club evaluation and previous research.

Pupils in Bug Club schools made significantly greater progress in their reading than children in schools not yet using Bug Club from baseline to A1 (two terms) and from baseline to A2 on decoding (three terms). However, the magnitude of the overall effects were small. At A1 and A2, pupils in the Bug Club schools made approaching two more points (1.75/1.92) progress on the standardised reading measure (es = .18 and .16 respectively) than pupils in control schools. The advice in conducting RCTs is to run a power analysis to estimate an appropriate sample size. The typically resulting large samples provide confidence in the significance of any differences observed between intervention and control groups, hence the statistical significance, but those differences may be very small. Significant differences between Bug Club and control children, using a waiting list design, were not found in the replication study, where pupils who participated in Bug Club over five months, from A2 to A3, made expected but not accelerated progress. Y1 Bug Club pupils recruited at the beginning of the study were followed up over five terms at the end of which time they were 11 months ahead on their expected InCAS age equivalent reading score, relative to their chronological age. At baseline they had been 1 month behind in terms of reading age. In this study there was a significant interaction effect between

Bug Club and social disadvantage (as measured by FSM and PP). Bug Club was most effective for schools with a high proportion of pupils on FSM at A1 and for pupils on PP at A2. The lack of consistency on whether this poverty effect was at school or pupil level leaves a question mark as to the precise locus of the effect and since the finding was not anticipated on any previous theoretical or empirical grounds would need replication. At A2, after three terms with Bug Club, children on PP had made an average gain of 9.4 points on the standardised reading measure, compared to 4.0 points by the equivalent control children, a small but useful effect (es = .34). However, these findings for pupils on PP were not repeated in the replication study.

How can we account for these findings? Firstly, The positive findings of Bug Club in the first year, particularly in decoding, might partly be explained by its systematic approach to phonics teaching in addition to an integration of decoding and comprehension skills (Higgins et al). No similar RCTs with this specific age group and of this type of reading programme currently exists that measures vocabulary, decoding and comprehension. However, compared to the closest; Open Court Reading and Reading Street, the ES for standardised reading was slightly higher than for Bug Club, whilst Bug Club was particularly effective for decoding.

Secondly, the success of Bug Club for children from poorer families, though tentative, might be explained, in part, by the home and school collaboration supported by the programme. The specific aspects of SES that contribute to lower outcomes are multifaceted. Low SES families often lack the resources to provide the range of leisure and social activities experienced by children from high SES families (Esping-Andersen et al. 2012). The home environment is less likely to provide the different types of opportunity that associated with supporting the development of reading, such as easily accessible reading material (Yeo, Ong and Ng 2014). The inclusion of texts in the Bug Club programme, which are accessible at home and school using different media and reading genres, and designed to promote pupil motivation and engagement, might be expected to be particularly valuable for socially disadvantaged children.

In addition, one of the three interventions identified as having the most impact for this group of pupils in a recent review; feedback and progress monitoring, is an integral part of the Bug Club programme though, for example guided reading and resources available online for home use (Dietrichson et al. 2017).

However, accounting for the findings from the second year of the study is more complex, namely; a) why did the replication study of two terms of Bug Club in the second year of the evaluation not elicit the significant gains reported in the first year and b) why did the accelerated progress evident from baseline to A1 flatten out in the second year of the study? One obvious explanation is that fidelity to the Bug Club programme was not consistent in the second year of the programme and the poor reporting on usage by teachers in the second year adds weight to this proposition. The fact that 13% of the variance on children's performance on the standardised reading measure was explained at school level is consistent with this explanation. As teachers' reports of their use of Bug Club from A2 to A3 were very incomplete we cannot make robust judgments about whether teachers implemented the programme in a different way in the second year of the study. However, in the second year of the Bug Club evaluation, a new primary school curriculum with changes in assessment and reporting system were introduced in England which might have influenced the time available for staff to sustain the same attention on the Bug Club programme. Moreover, the Bug Club programme is designed to be flexible is often used in conjunction with other materials, such as other learning websites, other reading schemes and self-made teacher materials. This flexibility is intended to give teachers the freedom to implement Bug Club in a way that will best engage their pupils and achieve learning outcomes specific to their school syllabus but presents a considerable challenge in an RCT context and other pedagogies and resources in the classroom that might affect reading attainment. With regard to why accelerated progress flattened out in the second year of the study, there are no studies similar to Bug Club for comparison that could inform us about the

expected gains longitudinally, though we might presume that it would be difficult to sustain accelerated progress indefinitely. It also raises questions about the desirability of variety year on year in the child's experience of the curriculum in terms of activities and materials, beyond the obvious need to be developmentally appropriate. Finally it is possible that teachers 'followed the recipe' in the first year but then adapted more in the second year, perhaps beginning to focus on other areas of concern.

Although the effect sizes found in this study were small, this is typical for RCTs in education. A recent analysis of 141 RCTs in education (1,222,024 students) conducted by the EEF in the United Kingdom and the National Centre for Education Evaluation and Regional Assistance (NCEE) in the United States, found a mean effect size of 0.06 standard deviations and ranged from -0.16 to 0.74 across the 141 studies (Lotie-Forgues and Inglis 2019). A previous study found an average effect size of 0.23 for non-randomised quasi-experimental studies and 0.16 for randomised experimental studies (Cheung and Slavin 2016). The effect sizes for Bug Club (0.16 and 0.18) are comparable with the findings from both these studies.

Limitations of the study

The present study had limitations and many of these reflect the challenges highlighted in the wider research on RCTs in education. Firstly, we do know that studies have found associations between fidelity to programme implementation and outcomes. A review of 500 studies clearly showed that the level of implementation achieved is an important determinant of programme outcomes and that achieving good implementation increases the chances of programme success in statistical terms (Durlak and DuPre 2008). Therefore, to fully understand and explain the results of the Bug Club evaluation and any other RCT, a robust and comprehensive process evaluation should accompany a RCT (Hanley, Chambers and Haslam 2016). Although it is interesting to note that the more recent, scaled up effectiveness trial of Open Court Reading reported high levels of implementation fidelity with no significant positive effects for the whole

cohort. There is still much to consider about how to conduct fidelity studies as it has been demonstrated that high fidelity scores can conceal significant teacher variation, which might have implications for pupil outcomes (Troyer 2017).

Secondly, the contribution of individual teachers to the results of any RCT is an important consideration. As with most RCTs in education, the Bug Club evaluation used standard approaches to ensure that there were no significant differences in characteristics between the pupils and schools in the study. But we have yet to identify how to 'control', if possible or ethical, for teacher characteristics in terms of their efficacy including quality (as opposed to years) of teaching experience. Mendive and colleagues (2016) in their recent RCT of a preschool teacher professional development programme in Chile, found that teaching practices that were 'novel' as opposed part of a regular feature of practice were less likely to be addressed, even if part of a reading programme (Mendive et al. 2016). In this study, teachers in the experimental and control groups spent more time on emerging writing rather than vocabulary. Might it be that the stronger finding for word decoding rather than reading comprehension in the Bug Club study was also a reflection of the education policy focus in England on the teaching of phonics over the past ten years? Hanley and colleagues (2016) raise the interesting issue of allowing teachers time to embed complex programmes or interventions into their practice before a RCT begins and thus moving away from immediate to more medium term effects of a programme. This does raise questions for traditional RCT methodology and the complexities associated with aligning classroom practices and traditional RCT approaches.

Implications for practice

How might professionals in schools use the finding to inform their planning and practice? The accelerated progress of pupils in receipt of PP in the Bug Club schools suggests that evidence

informed approaches that are designed for all children, can also support progress for children from families with low SES. This is important, as such a programme has the potential to prevent pupils who might have been at risk, from falling behind in reading. In this way, a whole school reading programme is an early intervention in itself with pupils avoiding potentially damaging and long lasting labels and without the need to draw on any additional resources. At the very least, it might be sensible to argue that where there is positive evidence of effectiveness and no negative evidence, a decision to adopt a programme would be warranted in circumstances where there are no clear implications for diverting substantial resources from elsewhere. With regards to effectiveness for a particular group of children, schools might be informed by their own strategic plans.

Teaching children to read effectively and with enjoyment, requires, in no small part, on skilled teachers using evidence informed pedagogy and curriculums. It is possible to achieve this in the early primary years without the use of a whole school reading programme. However, for some schools it will be attractive, for many reasons, to use a programme that has at least some evidence base and which might be used as a framework to inform the literacy programme throughout the school. These reasons might include, for example, ensuring consistency of approaches across classes in the same year group and as a response and safeguard against high teacher turnover and/or inexperienced literacy teachers.

To conclude, 'Evidence-based research and RCTs in education are in their infancy in the UK…' (Hanley et al. 2016, 296) and this the current study, as the first published RCT of a whole school reading programme undertaken with pupils in years 1 and 2 of primary school in the UK, makes a fundamental contribution to our emerging understanding of what pupil progress might be expected as a result of a reading programme and the challenges of conducting a RCT within a school context.

References

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Oxford:Psychology Press.

Borman, G, Dowling, NM., & Schneck, C. (2007). "The national randomized field trial of Open Court Reading." *Madison, WI: University of Wisconsin*.

CCEA. (2007). Northern Ireland Curriulum. http://www.legislation.gov.uk/nisr/2007/46/pdfs/nisr_20070046_en.pdf

Cheung, A. C. K., & Slavin, R. E. (2016). How Methodological Features Affect Effect Sizes in Education. *Educational Researcher*, *45*(5), 283–292. https://doi.org/10.3102/0013189X16656615

Department for Education. (2018). National curriculum assessment at key stage 2 in England, 2018 (provisional).

Department for Education. (2013). English programmes of study: key stages 1 and 2. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/335186/PRIMARY_national_curriculum_-_English_220714.pdf

Desjardins, R., Thorn, W., Schleicher, A., Quintini, G., Pellizzari, M., Kis, V., & Chung, J. E. (2013). *OECD skills outlook 2013: First results from the survey of adult skills*.

Dietrichson, J., Bøg, M., Filges, T., & Klint Jørgensen, A. (2017). "Academic interventions for elementary and middle school students with low socioeconomic status: A systematic review and meta-analysis." *Review of Educational Research* 87 (2):243-282.

Durlak, A., & DuPre, E. (2008). "Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation." *American journal of community psychology* 41 (3-4):327-350.

Esping-Andersen, G., Garfinkel, I., Han, W., Magnuson, K., Wagner, S., & Waldfogel. J., (2012). "Child care and school performance in Denmark and the United States." *Children and Youth Services Review* 34 (3):576-589.

Gross, J. (2006). *The long term costs of literacy difficulties*: KPMG Foundation.

Hanley, P., Chambers, B., & Haslam, J. (2016). "Reassessing RCTs as the 'gold standard': synergy not separatism in evaluation designs." *International Journal of Research & Method in Education* 39 (3):287-298.

Higgins, S., Martell, T., Waugh, D., Henderson, P., & Sharples, J. (2017). Improving literacy in key stage 2.

Kuczera, M., Field, S., & Windisch, H. C. (2016). Building skills for all: a review of England. *Policy Insights from the Survey of Adult Skills. Paris, France: OECD*.

Lortie-Forgues, H., & Inglis, M. (2019). Rigorous Large-Scale Educational RCTs Are Often Uninformative: Should We Be Concerned? *Educational Researcher*, *48*(3), 158–166. https://doi.org/10.3102/0013189X19832850

Mendive, S., Weiland, C., Yoshikawa, H., & Snow, C. (2016). "Opening the black box: Intervention fidelity in a randomized trial of a preschool teacher professional development program." *Journal of Educational Psychology* 108 (1):130.

Merrell, C., & Tymms, P. (2007). "Identifying reading problems with computer-adaptive assessments." *Journal of Computer Assisted Learning* 23 (1):27-35.

Pignone, M. P., & DeWalt, D. A. (2006). Literacy and health outcomes. *Journal of General Internal Medicine*, *21*(8), 896-897.

Scanlon, D. M., Anderson, K. L., & Sweeney, J. M. (2016). *Early intervention for reading difficulties: The interactive strategies approach*. Guilford Publications.

Sénéchal, M., & Young, L. (2008). The effect of family literacy interventions on children's acquisition of reading from kindergarten to grade 3: A meta-analytic review. *Review of Educational Research*, *78*(4), 880-907.

Slavin, E., Lake, C., Chambers, B., Cheung, A., & Davis, S. (2009). "Effective reading programs for the elementary grades: A best-evidence synthesis." *Review of Educational Research* 79 (4),1391-1466.

Troyer, M. (2017). "Teacher implementation of an adolescent reading intervention." *Teaching and Teacher Education* 65:21-33.

Vaden-Kiernan, M., Borman, G., Caverly, C., Bell, N., Sullivan, K., Ruiz de Castilla, V., Fleming, G., Rodriguez, D., Henry, C., Long, T., & Hughes Jones, D. (2018). "Findings From a Multiyear Scale-Up Effectiveness Trial of Open Court Reading." *Journal of Research on Educational Effectiveness* 11 (1):109-132. doi: 10.1080/19345747.2017.1342886.

Wilkerson, SB., Shannon, LC., & Herman, T. (2007). "The efficacy study of Pearson Education's Reading Street program: Year two." *Retrieved June* 5:2009.

Yeo, L., Ong, W., & Ng, C (2014). "The home literacy environment and preschool children's reading skills and interest." *Early Education and Development* 25 (6):791-814.

Table 1: Primary reading programme evaluations

Programme/ Study	Category (Slavin et al. 09)	Design	Length	No.	Grade	Measures	Vocab	Decoding	Comprehension	Overall effect size (Cohen's d)
Open Court Reading Borman et al. 07	Curricula	Random quasi-experiment	1 year	307	1	Terra Nova	0.22		0.06	0.17
Reading Street Wilkerson et al. 2007	Curricula	Random quasi-experiment	1 year	387	1	Gates MacGinitie			0.15	0.15

Table 2: Overview of different groups for each phase of data collection and analysis

Pupil testing points	Original Bug Club Schools: Groups		Waiting list control schools: Groups	
	Y1	Y2	Y1	Y2

Baseline (Jan 2015)	A	B	C	D
A1 (July 2015)	A	B	C	D
A2 (January 2016)	E (new Year 1 pupils)	A	F (new Year 1 pupils)	C
A3 (July 2016)	E (new Year 1 pupils)	A	F (new Year 1 pupils)	C

Table 3: Numbers of pupils in Years 1 and 2 BC and control schools at baseline, A2 and A3[1]

Sample	Condition	Year Group		Subtotal	Total
		Year 1	Year 2		
Baseline: Jan 2015	Control	359	377	736	
	BC	346	359	705	1441
A1: July 2015	Control	334	346	680	
	BC	302	291	593	1273
A2: Jan 2016	Control	316	376	692	
	BC	348	385	733	1425
A3: July 2016	Control (novice BC)	287	344	631	
	BC	303	336	639	1270

Table 4: Average weekly usage of materials in BC schools: B-A1

	Average weekly usage*

[1] At A3 control schools had been using BC since A2 and had therefore become novice BC schools.

	B – A1 (sd)
Use of print resources	2.21 (.65)
Use of online resources	1.98 (.60)
Contexts where print resources used	2.32 (.68)
Contexts where online resources used	1.89 (.60)
Contexts where phonics resources used	2.03 (.63)
Use of non BC materials	2.03 (.52)

* 4=daily; 3=2 to 3 times per week; 2=once per week; 1=not used in the week.

Table 5: Standard scores for reading and developed ability at baseline for BC and control groups (excluding children not also assessed at A1)

Sample	Variable	Condition	N	M	SD	*P*
Baseline	Developed ability	Control	652	97.64	17.51	-
		BC	569	96.31	17.51	
	Reading	Control	680	103.02	14.33	-
	Standardised	BC	593	103.80	15.09	

Table 6: Comparison of BC and control pupils' literacy progress from baseline to A1, baseline to A2 (Y2 at A2 only) and replication study A2 to A3

	Means & SDs A1		B-A1 intervention effects			
	Bug Club	**Control**	*B*	*SE*	*p*	**Effect size**
	(N=593)	**(N=680)**				**(Cohen's**
						d)
	M (SD)	M (SD)				

Reading Standardised	109.01 (15.18)	106.68 (14.78)	1.75	0.90	.050	.18
Subscales						
Word recognition	7.59 (1.90)	7.37 (1.80)	0.098	0.077	.207	.09
Word decoding	8.17 (2.15)	7.86 (2.11)	0.300	0.146	.039	.18
Comprehension	7.54 (1.86)	7.31 (1.81)	0.207	0.106	.050	.12
Spelling	7.54 (1.72)	7.30 (1.73)	0.252	0.107	.019	.23
Picture vocabulary	7.56 (2.05)	7.37 (2.01)	0.171	0.124	.170	.13
Chronological age	6.92 (.59)	6.92 (.57)				

	Means & SDs A2 (Y2 only)		**B-A2 effects**			
	Bug Club (N=318) M (SD)	**Control (N=313)** M (SD)	**Beta**	**R^2**	**t**	
Reading Standardised	108.87 (15.04)	106.54 (16.46)	.055	.003	2.12 *	.11
Subscales						
Word recognition	7.68 (1.74)	7.51 (1.78)	.023	.001	.89	
Word decoding	8.21 (1.96)	7.72 (2.08)	.119	.014	3.89 ***	.24
Comprehension	7.52 (1.81)	7.53 (1.81)	.003	0	.07	
Spelling	7.46 (1.67)	7.39 (1.65)	.018	0	.60	
Picture vocabulary	7.75 (1.86)	7.51 (1.92)	.073	.005	2.50 *	.15
Chronological age	6.92(.29)	6.95 (.29)				

	Means & SDs A3 (A1 for controls)		**A2-A3 effects**	

	Bug Club M (SD)	Control M (SD)	Beta	R^2	t	
Reading Standardised	106.55 (16.70)	106.90 (14.49)	.007	0	.45	
Subscales						
Word recognition	7.40 (2.00)	7.40 (1.78)	.013	0	.97	
Word decoding	7.71 (2.27)	7.91 (2.09)	-.006	0	-.29	
Comprehension	7.56 (1.92)	7.32 (1.80)	.030	0	1.51	
Spelling	7.31 (1.79)	7.32 (1.72)	.033	.001	1.91	
Picture vocabulary	7.58 (2.13)	7.41 (2.01)	.045	.002	2.60 **	.09
Chronological age	6.92 (.59)	6.92 (.57)				

DV=Dependent Variable; SD=standard deviation

*p<.05 **p<.01 **p<.001

Table 7: Interaction effects of PP (pupil level) and condition on Reading Standardised at A1.

Dependent variable = Reading Standardised at A1				
Block	Variable entered	Beta	R^2	t
One	Reading Standardised baseline	.821		52.76 ***
	PP pupil level	-.082	.710	-5.27 ***
Two	Reading Standardised baseline	.820		52.91 ***
	PP pupil level	-.080		-5.162 ***
	Condition (experimental or control)	.051	.712	3.348 ***
Three	Interaction (PP x Condition)	.064	.714	3.035 **

PP=Pupil Premium

*p<.05 **p<.01 **p<.001

Table 8: Interaction effects of PP (pupil level) and condition on Reading Standardised at A2 (Y1 at baseline to Y2 at A2 only).

Dependent variable = Reading Standardised at A2

Block	Variable entered	Beta	R^2	t
One	Reading Standardised baseline	.741		27.73 ***
	PP pupil level	-.065	.572	-2.43 *
Two	Reading Standardised baseline	.740		27.75 ***
	PP pupil level	-.062		-2.31 *
	Condition (experimental or control)	.052	.574	2.00 *
Three	Interaction (PP x Condition)	.088	.578	2.41 *

Table 9: Interaction effects of PP and condition on Reading Standardised comparing controls (groups C and D) baseline to A1 with novice BC (groups C and F) A2 to A3.

Dependent variable = Reading Standardised at A3 (at A1 for controls)

Block	Variable entered	Beta	R^2	t
One	Reading Standardised A2 (baseline for controls)	.801		49.18 ***
	PP pupil level	-.104	.599	-6.38 ***
Two	Reading Standardised A2 (baseline for controls)	.801		49.16 ***
	PP pupil level	-104		-6.37 ***
	Condition (experimental or control)	.002	.601	.118
Three	Interaction (PP x Condition)	.022	.601	.985

PP=Pupil Premium

*p<.05 **p<.01 **p<.001

Table 10: BC pupils (Group A) followed up from baseline to A3 on Reading Age Equivalent

Assessment point	Chronological age Mean	Time from Baseline Mean	Reading Age Equivalent Mean
Baseline n=279	5y 11m		5y 10m
A1 n=264	6y 5m	6 m	6y 11m
A2 n=284	6y 11m	1 y	7y 9m
A3 n=284	7y 5m	1 y 6 m	8y 4m

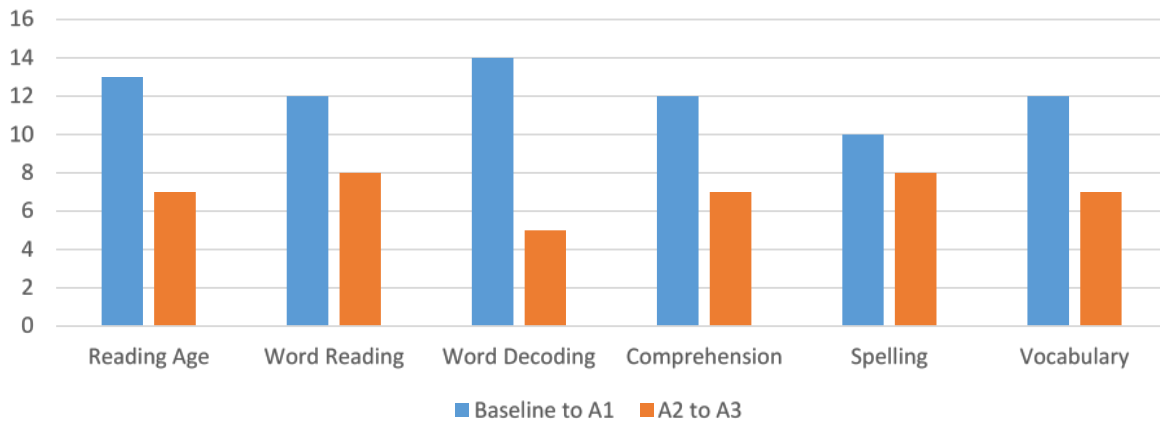PP=Pupil Premium

*p<.05 **p<.01 **p<.001

Figure 1: Mean gains in literacy from baseline to A1 compared with A2 to A3