Word and World Reading

Evaluation report and Executive summary

July 2015

Independent evaluators:

Beng Huat See, Stephen Gorard, Nadia Siddiqui, (Durham University)

The Education Endowment Foundation (EEF)

The Education Endowment Foundation (EEF) is an independent grant-making charity dedicated to breaking the link between family income and educational achievement, ensuring that children from all backgrounds can fulfil their potential and make the most of their talents.

The EEF aims to raise the attainment of children facing disadvantage by:

- Identifying promising educational innovations that address the needs of disadvantaged children in primary and secondary schools in England;

- Evaluating these innovations to extend and secure the evidence on what works and can be made to work at scale;

- Encouraging schools, government, charities, and others to apply evidence and adopt innovations found to be effective.

The EEF was established in 2011 by the Sutton Trust, as lead charity in partnership with Impetus Trust (now part of Impetus-The Private Equity Foundation) and received a founding £125m grant from the Department for Education.

Together, the EEF and Sutton Trust are the government-designated What Works Centre for improving education outcomes for school-aged children.

For more information about the EEF or this report please contact:

Robbie Coleman
Research and Communications Manager
Education Endowment Foundation
9th Floor, Millbank Tower
21-24 Millbank
SW1P 4QP

p: 020 7802 1679
e: robbie.coleman@eefoundation.org.uk
w: www.educationendowmentfoundation.org.uk

About the evaluator

The project was independently evaluated by a team from Durham University led by Professor Stephen Gorard.

Stephen Gorard is Professor of Education and Well-being, and Fellow of the Wolfson Research Institute at Durham University. He is a Methods Expert for the US government Institute of Education Science, member of the ESRC Grants Awarding Panel, and Academician of the Academy of Social Sciences.

His work concerns the robust evaluation of education as a lifelong process, focused on issues of equity and effectiveness. He regularly advises governments and other policy-makers, including oral and written advice to the House of Commons Education Committee every year since 2003. He is also a widely read and cited methodologist, involved in international and regional capacity-building activities, and used regularly as an adviser on the design of evaluations by central and local governments, NGOs and charities. He is currently an evaluator for the European Commission Directorate-General for Education and Culture, the Department of Work and Pensions, the Food Standards Agency, the Learning and Skills Information Service, and the Education Endowment Foundation. He is author of nearly 1,000 books and papers.

Contact details:

Professor Stephen Gorard
School of Education
Durham University
Leazes Road
DH1 1TA

p: 0191 334 8419
e: s.a.c.gorard@durham.ac.uk

Contents

Executive summary

The project

The Word and World Reading programme aimed to improve the reading comprehension and wider literacy skills of children aged 7–9 from low income families. The programme focused on improving the vocabulary and background knowledge (sometimes labelled 'core knowledge') of pupils, through the use of specially designed 'knowledge rich' reading material, vocabulary word lists, a read-aloud approach, and resources such as atlases and globes. The programme is based on the rationale that children need background knowledge to be able to comprehend what they read, and that improving background knowledge is an effective way to help struggling readers.

This pilot evaluation involved 17 primary schools from across England. Participating schools received training that emphasised the consistent and sequenced use of vocabulary, direct instruction, and teacher questioning. Year 3 and 4 classes in participating schools followed the approach for the whole 2013–14 academic year. The programme was developed and delivered by The Curriculum Centre, a charitable organisation which is part of Future Academies. The project was co-funded by the Greater London Authority (GLA).

The evaluation had three aims. First, to assess the feasibility of the approach and its reception by schools. Second, to assess the promise of the approach and provide recommendations that could be used to improve the approach in the future. Third, to provide recommendations that could be used to design any future trial, including an assessment of the appropriate size of any future trial.

Key conclusions
1. The Word and World Reading programme was introduced as intended, and was well received by the majority of primary schools participating in the project.
2. Some teachers felt that the programme had a positive impact on pupil learning, including improved vocabulary and writing skills.
3. In some lessons, teachers' subject knowledge did not appear to be sufficient to support an in-depth discussion with pupils about some of the topics within the programme curriculum. This suggests that additional training or support materials may have been beneficial.
4. The programme appeared to be more successful for older, higher attaining students, and less successful for Year 3 students or low attaining students. Greater differentiation, for example adapted vocabulary lists, may have made it easier for lower attaining students to engage with the programme.
5. The study did not seek to assess impact on attainment in a robust way, however the attainment data which was collected did not indicate a large positive effect. This suggests that any future trial of the programme should involve a large number of schools in order to provide a precise assessment of the cost-effectiveness of the programme. It may also be valuable to test the approach over a longer period of time.

What did the pilot find?

The pilot study found that the Word and World Reading programme was feasible and could be successfully integrated into the school curriculum. The programme was well received by most schools participating in the project. Five more schools registered interest in the project than were required by the original evaluation design.

There was mixed evidence of promise for the approach. Some teachers felt that the approach had led to noticeable improvements' in pupils writing and reading comprehension skills. In many lessons

pupils were highly engaged, sometimes to the surprise of teachers who believed that their classes would not enjoy the highly-structured nature of the programme.

Overall, pupils were positive about the approach, and appeared to enjoy the lessons more as the programme progressed. It appeared that Year 4 pupils were more engaged in lessons than Year 3 students. Some lower attaining students or students with English as an additional language appeared to find it harder to engage with the material, suggesting that some simpler texts should be provided.

In some lessons, teachers' subject knowledge did not appear to be sufficient to support a deep and engaging discussion of the material included in the curriculum. This might be improved by more intensive teacher training, and by pointing teachers towards additional resources or information sources that could be used to support lessons.

The study did not seek to assess impact on attainment in a robust way and the security of the quantitative findings was affected by a high level of attrition. However, the attainment data which was collected did not indicate a large positive effect. This suggests that any future trial of the programme should involve a large number of schools, in order to provide a precise assessment of the cost-effectiveness of the programme. It is also possible that a trial conducted over a longer time period would provide a better assessment of the true impact of the approach.

How was the pilot conducted?

Data was collected through lesson observations and interviews with pupils and staff by the independent evaluators, with input from developers. In total 12 visits were made to 8 participating schools and 56 classes were observed. In addition, eight visits were made to schools in both the intervention and comparison groups when pupils sat the end of project test.

Data was collected in participating schools and nine comparison schools in order to support an assessment of the appropriate size of any future trial. Schools were allocated to follow the approach or join the comparison group through random allocation. The impact assessment did not indicate that the approach had a large positive effect.

How much does it cost?

The cost is estimated as £50 per pupil. This includes the cost of globes, atlases, pupil workbooks, teacher workbooks, teacher supply cover when on training, teacher training, and other administrative costs. These represent an initial investment only, since the teaching resources can be re-used. The costs for subsequent implementation would be largely for replacement of staff and resources, and further professional development.

Question	Finding	Comment
Was the approach feasible?	Yes	All schools completed the project and the programme was well received.
Is there evidence of promise?	Mixed	Some teachers felt the programme had a positive impact on learning, but the quality of lessons delivered within the programme was variable.
Is the approach ready for a full trial without further development?	No	Prior to any future trial further development of the approach would be valuable. To detect an effect of a similar magnitude to the indicative effect detected in the pilot, a very large trial would be required.

Introduction

Intervention

The project evaluated is a year-long pilot trial of the Word and World Reading (WWR) programme. The rationale behind The Curriculum Centre's project is that children need background knowledge and understanding to be able to comprehend what they read, and those with a broad base of factual knowledge can build on this knowledge to enhance further learning. The programme aimed to address the vocabulary gap between disadvantaged and affluent children, and to enhance the literacy and comprehension abilities of pupils in early Key Stage 2 (Years 3 and 4). The intervention involves using specially developed knowledge-rich reading material, subject-specific resources, and vocabulary word lists. In addition, pedagogical continuing professional development (CPD) for teachers in history and geography is provided. The associated teacher training emphasises consistent and sequenced use of vocabulary, direct instruction, and teacher questioning.

Background

The Curriculum Centre's WWR programme had not previously been tested in English schools or elsewhere. Some aspects of the programme, however—including the importance placed on background vocabulary—are similar to other programmes that have been tested in the United States.

The most widely adopted scheme is the Core Knowledge Language Arts (CKLA) programme that was developed from the work of E. D. Hirsch on cultural literacy. Hirsch observed that some groups of students were able to understand passages of texts more easily than others, and that this systematic difference was due to lack of familiarity with the context (Hirsch, 1987). He believed that this was because the curriculum in the early years did not allow children to build on their basic background knowledge. Consequently, he and his foundation developed a reading programme that included both phonic drills and background knowledge ('core knowledge') that aimed to help develop students' comprehension skills. It was first piloted in 1990 in a Florida elementary school. Since then it has gone through several revisions based on feedback from schools, and has been implemented in thousands of schools across the US.

To date, the biggest evaluation of the CKLA programme was conducted by Johns Hopkins University (Stringfield et al., 2000): this was a national evaluation of 12 schools following the Core Knowledge approach ('CK') in seven states across the US. Two cohorts of pupils—those who started CK in the first grade or age 6 (n = 663) and those who started in the third grade or age 8 (n = 653)—were tracked through to third grade (Cohort 1) or fifth grade (Cohort 2). At the end of the third year, only data from 61% and 65% of Cohort 1 and Cohort 3 respectively was available for analysis. Data was analysed from four matched pairs of comparator and CK schools in Florida, Maryland, Texas and Washington. Results from the norm-referenced tests suggested that CK had an overall negative impact on the reading of children in Cohort 1 (ES = -0.06). All states apart from Florida registered a negative impact on reading (pp 54–57). The authors explained that the exceptionally poor performance of CK pupils in Maryland (where it was argued that CK had been poorly implemented) had skewed the overall results. The effects on low achieving pupils appeared more promising (ES = 0.25). CKLA was also found to have a negative effect on the reading of children in Cohort 2 (ES = -0.08). The intervention showed a greater negative effect for low achieving pupils (ES = -0.53). This was again put down to the poor performance of CKLA pupils in Maryland. The conflicting results for low achieving pupils in Cohort 1 and Cohort 2 suggest that the security of the findings is questionable.

In a preliminary evaluation, Dartnow et al. (2000) assessed the effects of the CK programme in its first two years of implementation in eight schools (four comparison and four CK schools) across the US (Maryland, Texas, Florida and Washington). The study used a matched comparison of demographically similar schools as controls. Results for the first grade cohort showed that only one

site (Florida) reported a positive impact on overall and subtest analyses (average ES of 0.61 for reading and 0.40 for maths). In the other three sites (Maryland, Texas and Washington), no differences were found for overall or sub test scores. Results were more mixed for the third grade cohort. As with the first grade cohort, only one site (Washington) reported a positive impact on the overall score and subtest analyses of reading skills (ES of 0.31). A negative impact was reported in Maryland for reading skills (ES = -1.31). In Florida and Texas, no differences were found between CK group and comparison group for overall scores. Results from a bespoke test created by the developer showed a positive impact in only two sites for both cohorts. The authors explained that the possible reason for the mixed results could be the varying degree of implementation among the three sites and the demographic composition of pupils. In Maryland, for example, it was noted that there were lower levels of implementation compared to the other sites. There were also proportionately more African Americans (98%) in the Maryland site than in Florida and Washington: in these latter sites the majority of pupils were white—82% and 70% respectively.

A more detailed report of the Maryland school where implementation was low suggested a very high attrition rate in all of the schools at the end of the third year (Iver et al., 2000). Cohort 1 was tracked from first grade to fifth grade, while Cohort 2 was followed-up for three years (from third grade to fifth grade). From an initial total of 1,207 pupils in the first year, only 59% of the experimental group and 58% of control pupils remained at the end of the third year. By year five, the numbers dropped further to only 41% and 40% respectively. The study started with six pilot schools and six matched comparison schools, however in the second year, one CK school and its paired control school were dropped, and in the fourth year another pair was dropped when the CK school abandoned the programme. Academic outcomes were measured using the Comprehensive Test of Basic Skills (CBTS) and the Maryland School Performance Assessment Program (MSPAP). Analysis of the first grade cohort at the end of the third year showed that all CK schools, apart from one, made greater gains than their control schools. The one low performing CK school did so badly compared to the control that it skewed the overall results. Overall, CK schools appeared to perform more poorly than control schools. Average Normal Curve Equivalent (NCE) gains in CBTS Reading Comprehension were 4.8 for CK schools and 6.4 for control schools. Excluding the low implementation school from analysis, CK schools made bigger gains (average NCE gains = 8.1) than control schools (NCE gains = 4.2). Analysis using MSPAP showed mixed results with some CK schools making bigger gains than control schools, while others did not. However, when the gains in NCE scores were tracked over time, it was observed that there was only a substantial gain score in one of the schools, between the first and the second year of implementation. In Schools A, C, D and E, the control group outperformed intervention group (pp 37–41). In subsequent years there was no clear evidence that intervention schools were progressing faster than control schools. Even though the researchers argued that this was because School E was a low implementer school, it can also be argued that if CK had not been implemented at all, then intervention pupils should have been similar to control pupils. The fact that their scores dropped compared to control pupils may suggest that the little implementation they had was more harmful than no implementation at all.

In an earlier study by Whitehurst-Hall (1999) mixed results were also reported. This was also a longitudinal study following children over three years using a matched comparison design. In this study 301 Seventh and Eighth Grade pupils in one state in the US were assessed on reading, language and maths using the Iowa Test of Basic Skills (ITBS) subtests. Positive impacts were noted for some measures but not others. For example, on tests of comprehension, CK pupils outperformed control (ES = +0.07), but not on tests of language and expression, or sub-tests of spelling and punctuation (ES = 0.001). The author also reported no 'statistically significant' differences at the 0.05 level between CK and control pupils in terms of grade retention (where pupils are required to repeat a year) and grade failure.

An independent evaluation of the Core Knowledge curriculum conducted across schools in Oklahoma City in the US seemed to suggest a positive impact of CK on a range of subjects (Core Knowledge

Foundation, 2000). CK pupils reportedly obtained higher scores than non-CK pupils in reading comprehension (58.1 vs 55.1), vocabulary (59.8 vs 55.3), and social studies (58.3 vs 53.4). This was a matched comparison study where students taught the CK curriculum were randomly matched (using computer) with pupils not taught the curriculum on seven variables, including pre-test score on Iowa Test of Basic Skills (ITBS), sex, ethnicity and free lunch eligibility. Pupils were assessed on both the Norm-Referenced ITBS and the Oklahoma Criterion-Referenced Tests. On both tests, CK pupils performed, on average, better than those in the comparison group. Results were also promising when comparisons were made with matched students from the previous year using ITBS scores. CK pupils again obtained higher scores than their matched non-CK peers in reading comprehension (57.6 vs. 53.1), reading vocabulary (58.8 vs. 54.7) and social studies (60.4 vs 56.0). The matched, rather than randomised, design of the study may weaken the security of the findings. In addition, the full report was not available at the time of publication, which means that it is not possible to assess the level of attrition in the study.

Evaluation of the Core Knowledge early literacy pilot in New York City in 2012 (The NYC Core Knowledge Early Literacy Pilot 2012) reported promising gains in reading tests (Woodcock-Johnson III) especially in kindergarten, although the differences decreased by the third year. Using the standardised TerraNova test, however, no differences were detected in oral reading comprehension and vocabulary. Children who were new to the programme made the most gains. Although the report suggested that children who were on the programme for the longest had the highest post-test scores compared to those with only one and two years of exposure, it has to be mentioned that these children had higher pre-test scores too (i.e. they started on a higher base).

Although widely implemented, the evidence base linking the CK approach to improved literacy is currently underdeveloped. Evaluations to date have commonly adopted matched designs and have been developer led or funded. It would therefore be valuable to conduct further independent evaluations of the approach, preferably using randomised designs. The WWR approach has similarities with CK, and also draws on the work of other academics such as Walter Kintsch. This new evaluation builds on previous work in this area, and is the first study of this type of approach in English schools.

Objectives

The aim of this pilot trial was to look at the feasibility of the intervention, suggest formative improvements in delivery, and to estimate the impact 'effect' size—necessary to determine the appropriate size of any future evaluations. The objectives were:

- to test the feasibility of running the programme in a range of schools for a large-scale effectiveness trial;
- to provide a formative assessment of the protocol, and delivery, particularly through the views of staff and pupils;
- to describe the fidelity of implementation of WWR through process evaluation; and
- to estimate the likely magnitude of the effect of the Word and World Reading pilot on the reading comprehension of Year 3 and 4 pupils in primary schools with a high proportion of disadvantaged children in order to inform a future trial.

Project team

The programme was developed by The Curriculum Centre (TCC). TCC was responsible for developing the tools and instructional materials, training teachers, and supporting and monitoring the implementation of the programme in the classroom. The Curriculum Centre is a charitable organisation whose core mission to equip young people with the knowledge and skills to thrive in the 21st century. It is part of Future Academies, a multi-academy trust that operates academies and runs

school-centred initial teacher training in Westminster, London. It offers a range of content-rich curriculum from primary through to secondary schools. TCC conducted the training of teachers, subsequent monitoring, and support of programme delivery in schools. The evaluation of the programme was carried out by independent evaluators from Durham University.

Ethical review

Ethical approval was sought and given by Durham University Ethics Committee. Data was managed in accordance with the Data Protection Act (1998). The evaluation was conducted in accordance with the British Educational Research Association ethical guidelines in that all pupil data was treated in confidence. Only the team of researchers working on the project had access to the data. All the data submitted from schools was screened for missing data or double entries. Results are reported in aggregate: no schools or individuals can be identified.

Methodology

Process evaluation methods

The process evaluation was set up to provide formative evidence on all phases and aspects of the intervention—from the selection and retention of schools, through initial training and conduct of the intervention, to evaluating the outcomes. The evidence would be used to assess fidelity to treatment, the perceptions of participants—including any resentment or resistance—and to advise on improvements and issues for any future scaling up.

For this project, the process evaluation was conducted by the independent evaluators, with input from the developers. The latter conducted the recruitment of schools and training of teachers, monitored the delivery of the intervention, and collected formal records and the views of the staff. As part of the intervention, one of the programme developers visited schools and monitored the delivery of the programme. One visit was made per school per term. Each visit was preceded by a discussion with the member of staff who was in charge of the programme in the school. Discussion centred on programme-related issues, including perceptions of the programme and the resources. At the end of each visit the developer held a one to one debriefing session with the observed teachers to give them feedback on what had gone well in the lesson and what had not, and how the lesson could be improved. Similarly, teachers were encouraged to give an account of their own lesson, focusing on four main points: teacher delivery, pupil response, the strengths of the programme, and recommendations for improvement. A Lesson Visit Form was completed documenting the method of delivery, questioning and feedback styles used by teachers, and the work completed by pupils in their workbooks.

Three of the evaluators attended the staff training sessions as participant observers during which they talked to participants and the project deliverers to get their perceptions of the programme. There was no formal observation schedule, but evaluators noted the content and delivery of the training, and teachers' reaction.

The fieldwork, which aimed to assess the fidelity of implementation, formed the bulk of the process evaluation and included classroom observations as well as interviews with pupils and staff: it aimed to determine whether anomalies in pupil progress were due to the programme itself, or to implementation issues. We also wanted to note if the programme had changed teachers' teaching behaviour or children's engagement with literacy activity. Did the children enjoy the lessons? Did the programme lead to discernible differences in writing and reading skills?

Classroom observation visits to schools were arranged with a member of the Curriculum Centre staff. These were carried out once at the beginning of the intervention to observe the delivery of the programme—noting inconsistencies or any departures from the programme protocol—and also to note pupils' reaction to the programme and teachers' ability to use the resources. Another round of visits was carried out towards the end of the intervention, this time looking for changes in teacher behaviour and any differences in children's learning.

In total, 12 visits were made to 8 treatment schools to observe the process of implementation in 56 classes. In addition, 8 visits were made (3 control and 5 intervention schools) to observe the administration of the post-test to ensure the testing was carried out fairly and consistently across schools. As this was the first time that the long version of the PiE test was used, evaluators also took the opportunity to look at pupil and staff reaction to the test.

During these visits, both pre-arranged and ad hoc interviews were conducted with staff and pupils to understand their reaction to the intervention, and how the programme could be improved. Arrangements were made via the lead programme developer. These interviews were loosely

structured but focused on assessing teacher and pupil perceptions of the programme—what they thought had contributed (or would contribute) to the success of the programme, and the barriers to effective delivery. Pupils were also asked what they liked, or did not like, about the intervention.

The observations and focus group interviews were intentionally kept very open and unstructured so that observers and interviewers would not be constrained by formal procedures that might bias their work. The intention was to let the interviewees focus or talk about things that mattered to them.

Assessments of the quality, relevance and utilisation of teaching resources (teacher handbooks, pupil workbooks, pictorial and video images, globes and atlases), and the perception of both staff and pupils to these materials, were also part of the the process evaluation.

Trial Design

To estimate the likely magnitude of effect of the Word and World Reading programme on the reading comprehension of Year 3 and 4 pupils in primary schools with a high proportion of disadvantaged children, the pilot used a standard two-group waitlist design. A total of 17 schools were recruited from diverse geographical regions to include a range of schools of different sizes and levels of disadvantage. Nine were randomised to receive the intervention over one school year. The other eight formed the control group and followed regular practice for the year. To reduce potential post-allocation demoralisation, it was agreed with EEF and TCC that control schools would join a 'waitlist' and receive the resources and training after the trial was completed. In addition, a monetary incentive of £500 per school was provided to fund cover teachers to support teachers' attendance at training..

Schools, rather than classes, were allocated to conditions in order to minimise diffusion, and, at the request of the developers, to reduce the inconvenience for time-tabling, teachers and developer training.

Eligibility

Targeted schools were those from areas of high social deprivation or coastal deprivation, inner city schools with 'high challenge' pupils, and schools serving predominantly white working class communities. A range of large (five-form entry) and small (one-form entry) schools were also targeted for greater diversity. The schools recruited were not already involved in other similar programmes, to minimise the possibility of cross-contamination. All pupils in Years 3 and 4 were eligible for the trial.

Description of intervention

The intervention was administered to Year 3 and 4 children in schools that were randomly selected to follow the programme. Children in the control schools carried on their literacy lessons as they would normally do.

Word and World Reading programme is a whole class intervention carried out twice per week. Each lesson lasts 45 minutes. For this pilot the lessons were based on geography and history texts. Schools varied in how they arranged the lessons: some had geography and history every week, some had geography and history on alternate weeks, while others had geography in one term and then history the next. Generally, there was one comprehension exercise per week for each subject. Altogether there were 34 geography and 35 history planned lessons. The lessons were very structured, following a set sequence:

Part 1 (30–35 minutes)

- Each lesson introduces a new or related topic. The lesson begins with a required reading passage, no more than a page long, on the topic. The teacher reads the passage in the textbook aloud while pupils track the text with their fingers.
- After each paragraph the teacher pauses and asks questions about the text. Pupils may volunteer to answer or the teacher may call someone out or set up pair work. Pupils are encouraged to answer in full sentences.
- Through a series of questions (suggested in the teachers' handbook), teachers involve pupils in a general discussion making use of their prior knowledge. Teachers generally lead the direction of the lesson, but pupils may also contribute with their prior knowledge and experiences. Pupils are expected to answer in full sentences both orally and in writing.
- Images in the form of pictures are used to reinforce the concepts/words taught. These are either put up on the board as still images or presented as PowerPoint slides or films. Teachers are encouraged to supplement these resources with their own. Key words and key concepts are regularly repeated. A number of new words are taught explicitly in every session.

Part 2 (10–15 minutes)

- To consolidate the learning of these keywords, students are given a few tasks in their workbook. These consist of three short and clear mastery questions followed by a keyword exercise. For higher attaining pupils, there is a further optional extension question, mostly inferential, at the end of each chapter.
- Another key aspect of the intervention is instant feedback: teachers are expected to circulate and check pupils' answers and mark their workbook, giving immediate feedback and written suggestions for improvement. As part of the approach, pupils are required to use full sentences. The emphasis is on acquisition of knowledge and using the keywords correctly.

Resources

The teaching resources include text books for pupils, teacher handbooks, globes and atlases.

The pupil textbook is organised in units: each unit consists of a sequence of passages that are linked allowing pupils to build their conceptual understanding and vocabulary. For example, the history text starts with a passage on civilisation and how it started, followed by a passage about farming and the Mesopotamian civilization. The geography text starts with a topic on the planet Earth, with passages about maps, oceans and rivers. Each passage is no more than a page. The texts are written in a simple language appropriate for the age group. The textbook is also the workbook: each passage is accompanied by some comprehension questions and keyword activities. In every lesson a small number of keywords are introduced. The keyword activities often require pupils to use the keyword in a different context to the one in which it is introduced to encourage mastery of the keyword. Each of these keywords are then included, where appropriate, in future lessons.

The teacher handbook has similar texts to the pupil textbook, but there are guidance notes for teachers and prompts indicating where to stop and repeat the information for students, or when to engage them in discussion. Suggested questions are also included. There are also some suggestions on the use of images provided.

PowerPoint slides are provided as supplementary materials for teachers to use as illustrations for each lesson. There are no images in the teacher handbook or pupil workbook.

Globes and atlases

Each treatment school is also supplied with globes (one globe per teacher and seven per class) and atlases (31 per class).

Testing

Children took the pre-test when they were in Years 2 and 3 before participation in the programme in Years 3 and 4 respectively. The pre-test was the fully standardised digital version of Progress in English (PiE 7 for Year 2, PiE 8 for Year 3), a non-adaptive test supplied by GL Assessment. This was the short form online test which comprised four exercises:

- Exercise 1: Spelling;
- Exercise 2: Grammar;
- Exercise 3: Reading comprehension (narrative); and
- Exercise 4: Reading comprehension (non-narrative).

The post-test was the PiE long version (which is only available in paper form). Year 3 pupils took the PiE 8 and Year 4 pupils, the PiE 9. The first four sections were similar to the adaptive short version, but with four additional exercises: two were extended reading comprehension exercises (one narrative and another non-narrative), and one short and one long writing composition. The rationale for using the long version of the test was to assess pupils' writing ability—a primary aim of Word and World Reading programme. This was a change to the original protocol (that involved the short form again) requested by the developers and agreed by the funders after the evaluation had started.

Outcomes

The primary outcome measure was pupil progress from pre- to post-test in the Progress in English (PiE) test supplied by GL Assessment (as noted above). Pupil performance in the post-test was also calculated. Pre-test comparisons were made to check for group balance.

A bespoke test designed by the Curriculum Centre was suggested in the protocol, but this was not included in the analysis here because the test was deemed invalid by the evaluators as the content of the test covered materials that were taught explicitly to intervention children.

Other relevant data on pupil background characteristics including age, date of birth, sex, ethnicity, first language, special educational needs (SEN), and free school meals eligibility (FSM) were collected as part of the pre-testing. This data was uploaded at the outset for all pupils to the GL test system from each school's SIMS (School Information Management System), or similar, and then linked via UPN (Unique Pupil Number) to the individual post-test scores. This enabled secondary outcomes to be calculated in terms of progress scores for different sub-groups, especially FSM-eligible pupils only.

Both the pre-test and post-test were administered in collaboration with the project team. The pre-test was administered blind as it preceded knowledge of randomisation. Evaluators made sample visits to both intervention and control schools to observe the conduct of the pre-test. For the post-test, eight observation visits were made (five to WWR schools and three to control schools) to observe testing since schools now knew their allocation and such knowledge had the potential to influence teachers' behaviour in the way they conducted the test.

For the post-tests, the short-form exercises—containing two themed passages (fiction and non-fiction), spelling, and grammar—were marked by GL, the test provider. As GL did not have the capacity to mark the long form exercises (open response comprehension and writing tasks), this was undertaken by the evaluators. Detailed marking keys (from the test developer, GL) that have been rigorously moderated to guide teachers were provided for each evaluator marker. To ensure inter-rater reliability, ten sample scripts (each for PiE 8 and PiE 9) were picked for moderation. All markers marked the same ten scripts independently before coming together to compare their scoring.

Sample size

This pilot study aimed to offer the programme to eight primary schools, with a further eight schools in the same areas acting as control. Two year-groups in each school (Year 3 and Year 4) were targeted. The size of the sample was decided based on the assumption that there would be 90 pupils in each school (1.5 classes in each year group, and an average of 30 pupils per class). This meant that there would be 720 pupils (90 pupils each in eight schools) in each intervention group. A traditional 'power' analysis for a clustered sample makes a number of assumptions that are not valid here, but for illustration: if the intervention effect size for reading comprehension were around +0.3, then Lehr's approximation would suggest a minimum sample of around 178 cases per randomised group (Gorard, 2013). In reality the situation would be both better than this (because prior attainment scores would be available and included in the analysis) and slightly worse again (because the allocation to groups was by schools not individual pupils), nevertheless, a total sample of 1,440 is clearly in excess of the minimum needed to provide an estimated effect size from this formative evaluation. The result would provide a good estimate of the accuracy of the estimated effect size which could, in turn, be used to assess the sample size for any subsequent efficacy trial.

Recruitment

Schools were recruited through TCC contacts from a range of geographical areas. Targeted schools were sent an introductory information letter on the WWR pilot, outlining its aims and the conditions of participation. Schools first indicated their interest in the WWR pilot by verbal agreement. Following this, the schools then signed a non-binding contractual agreement with The Curriculum Centre. A Memorandum of Understanding was signed by each participating school to agree to being randomised to either treatment or control and to co-operate with the evaluators in providing the necessary data required for the evaluation. As the intervention was carried out as part of the school curriculum, it was no different to what the school would have done anyway (schools were going to implement the intervention whether it was being evaluated or not). Schools were responsible for parental opt-out consent to the evaluation.

It is not clear exactly how many schools, if any, were approached by the developers beyond those that ultimately agreed to participate in the project.

Randomisation

Schools were informed of the outcome of randomisation after the pre-test. Prior to the pre-test, schools uploaded pupils' background information onto the GL Assessment system. This included the name of the school and its URN (Unique Reference Number), pupil's surname and forename, UPNs, date of birth, sex, ethnicity, FSM eligibility, and first language. The schools with pre-test scores were randomised by the evaluators to either receive the intervention immediately or to the control (to receive the intervention a year later). To ensure blinding and to minimise bias due to knowledge of group allocation, evaluators carried out the randomisation using a mechanically shuffled pack of cards to select the treatment or waiting school, revealing the result only after the pre-test for all schools. There was one card for each school plus one (because of the odd number of schools), and an equal number of red and black signifying treatment or control. The last card was red, so the treatment group had one school more than the control group. Nine schools were randomised to receive the treatment and eight schools formed the control, continuing with 'business as usual'.

Analysis

An intention-to-treat analysis was used in this evaluation meaning that all pupils randomised to receive the intervention were tested and included in the analysis regardless of the actual amount of time spent on the intervention. For this reason, pupils who left their schools during the year were

tracked to their new destination as far as possible, so that their outcome data could be included in the analyses. The impact of the trial was estimated using the effect size (Hedges' g, as traditionally defined) of the difference between the treatment and control pupils in their gain scores between pre- and post-tests. The same calculation was repeated, for comparison, using the post-test scores in GL Assessment's PiE (long version) only. Sub-group analyses of pupils eligible for FSM, those with English as additional language (EAL), boys, and pupils identified with SEN were also conducted.

Two multivariate regression analyses were conducted using gain scores and post-test scores as the dependent variables. The predictors were prior test scores and pupil background characteristics, including sex, ethnicity, date of birth, first language, SEN, and FSM. The binary variable representing allocation to treatment or control was entered in a second step.

Readers may wish to note that significance tests and confidence intervals are not presented in this report. These do not work as intended (Carver, 1978), are almost always misinterpreted (Watts, 1991), and can lead to serious mistakes (Falk and Greenbaum, 1995). Above all, they take no account of sample quality or attrition (Lipsey et al., 2012), being predicated on complete random samples of a kind never encountered in real-life research (Berk and Freedman, 2001). This kind of explanation for *not* providing p-value and the like should no longer be necessary in a report such as this; rather those who still use such approaches should be asked to account for themselves, perhaps most simply by explaining what their cited probabilities could possibly relate to.

The cost of the intervention per pupil was estimated based on the assumption of 30 pupils per class. This included the cost of teaching resources (textbook, teacher handbook, atlases, globes), teacher training, supply teacher cover while teachers were receiving training, and administrative costs.

Process evaluation

Training

Training was offered at individual treatment schools as well as at The Curriculum Centre's London office. Schools appreciated having the option of on site or off site training. Of the nine treatment schools, three schools sent their staff to the London office for training and six schools requested in-school training. In some schools where the majority of teachers initially trained had left, the trainer went in to schools to retrain the teachers.

Five evaluators attended three training sessions conducted for different schools (or a combination). The training began with the theories underpinning the Word and World Reading programme before moving on to the practical application of the theories. The aims of the programme and the main principles were explained to the teachers. The training also introduced teachers to the resources. One of the trainers then modelled a lesson and the participants got to practise some of the strategies suggested. All lessons, it was specified, were to start with introducing or reactivating prior knowledge, followed by reading aloud (by the teacher), pausing for questions after each paragraph. The main idea was the repetition of words and concepts. Throughout the session the trainers took questions from the teachers, most of which concerned theoretical points and how they related to instructions in the classroom.

Teachers questioned how lessons could be differentiated for children with different levels of current attainment; they also wanted to know how the exercises could develop spelling skills. In general, teachers seemed to like the intervention (as they were keen to improve reading scores) and seemed to be happy with the handbooks, the concept, and the tasks, as well as the additional resources like the globe and the atlas they would receive. At the end of the training teachers appeared excited and were actively planning how they would implement the programme, considering, for example how they might apply the tasks with EAL pupils.

The programme developer reported very good feedback from teachers on the training they received. The participants were particularly inspired by the theoretical aspect of the WWR programme. Many teachers commented on how the theories presented (particularly around a knowledge-led curriculum and the science of how pupils learn) resounded with their beliefs in education and the changes that needed to be made to education to make it more beneficial to pupils. They found the model lesson clear and useful. At the end of the training, participant teachers shared their excitement about the pilot and were eagerly looking forward to delivering the WWR programme in their schools.

Lesson implementation

Altogether, 56 lessons were observed. Lessons were generally well conducted, adhering to the structure suggested in the Word and World Reading pilot programme. Teachers were engaging and pupils were mainly enthusiastic and focused. Some lessons were more inspiring than others. On the whole, Year 4 lessons appeared more successful than Year 3 lessons where teachers seemed to be struggling to generate enthusiasm among the pupils. However, having said that, pupils were largely engaged, asking pertinent questions, and on task. In most of the lessons observed, students were focused at all times and the lessons flowed smoothly. There was a natural progression from one stage of the lesson to the next.

Teachers generally followed the WWR sequence, starting the lesson with a recapitulation of the topic, followed by oral questioning, reading of text in short chunks followed by oral questioning, and comprehension and vocabulary exercises. As suggested by the programme guidelines, all the teachers adopted a combination of strategies in oral questioning. Pupils were asked to answer questions either individually or as a class. Pupils also had the opportunity to talk to their partners.

Almost all the teachers observed insisted on pupils answering in full sentences both orally and in their written work. There were plenty of hands-on activities in the geography lessons where pupils got to look at their atlas and to touch and feel the globe. Teachers walked round to check pupils' answers as they worked on their comprehension exercises, giving instant feedback. Where pupils had wrong answers, this was circled and the correct answers were inserted. In a few cases, marking was not rigorous and mistakes were not spotted or corrected.

It appeared that the highly prescriptive and structured lessons were both an advantage and a disadvantage. Most teachers said they liked the fact that the lessons were planned for them and there was minimal preparation on their part; some, however, adhered so closely to the prescribed programme that the lessons appeared contrived and there was little opportunity for open discussions. In contrast, where teachers attempted to initiate discussions, their lack of general knowledge and confidence in taking the discussions beyond the text was sometimes apparent. Evaluators noted some factual errors during the observation visits, such as giving pupils wrong information about the topic.

All the teachers made use of the visual aids provided by the programme developers. Where teachers had supplemented these with their own materials, these were used effectively. Pupils were taught to remember the names of the continents using mnemonics. In some classes, pupils struggled to remember the mnemonics. In a history lesson the teacher showed a video clip of the pyramids in Egypt and the inside of the pyramids. The teacher made his own 'sentence starter' slides to prompt pupils to answer questions in full sentences. This trained them to use the question as a starting point. For example:

'A monument is...'

'Khufu wanted to build a huge pyramid because...'

The focus on sentence construction was also present in other lessons. In a Year 3 geography class pupils practised reciting the correct answers to the questions by putting a part of the comprehension question and a part of the written text together, verbally, before writing them down. The teacher constantly emphasised the need for full sentence completion, for example saying: 'I'm looking for that nice full sentence again', 'Let's all say it again', reinforcing the precise answer.

There were instances where slight departures from the suggested WWR pilot structure were observed. For example, one teacher did not insist on full sentence answers in oral responses. Another teacher (who had not received the intervention training) did not follow the instructions by appointing individual pupils to read rather than reading aloud himself. Another teacher mistakenly thought that the vocabulary exercise was optional, so a number of pupils did not complete this section of the workbook.

In some of the pupils' workbooks there were grammatical and spelling mistakes that were not picked up by the teacher. One of the requirements of the WWR programme was that teachers should point out mistakes to the pupils by circling the mistakes in their workbook, but this was not consistently done.

Teaching resources

Slides

A common complaint among teachers was the lack of labelling for the pictures provided. These were not numbered and teachers reported difficulties in matching the pictures with the lesson plan. The quality of some of the images also needed improvement.

The lack of visuals was apparent in the geography class where there were plenty of opportunities to use them. The lesson on the rainforest was a good example where more visuals could have been used. At the same time, however, some teachers said that they liked the opportunity to make the lesson their own by making their own slides—one Year 3 teacher, for example, made slides of grapes and olives for a history lesson on Greece, and a Year 4 history lesson was enriched with photos and also little splices of documentary video about the pyramids. In this case, the children were genuinely engaged in the topic—why and how the Egyptians built the pyramids. The visual aids stimulated a lot of discussion among pupils with them asking, for example, whether the pyramids were still around today, and whether, if they went to Egypt, would they be able to see them; they also asked if one could actually go inside a pyramid and were visibly excited when told that they could. A teacher said that he found the geography lessons (with their maps and globes) 'more hands on' and had tried to develop ways to make his history lessons the same.

Globes and atlases

Both teachers and pupils thought the globes and atlases were beneficial. Every child had an atlas, and there was a globe for every five children. Children talked about how they liked looking at the globe and touching the 3-D effects on it. The colourful pictures in the atlas were also a talking point, however some teachers relied on these too much—there were lessons where these pictures were not sufficient.

Pupils' workbooks

The pupils' workbooks were not well-bound and often came apart.

A common suggestion made by teachers and pupils was to have coloured images in the textbooks. Pupils and teachers would like the images that come with each lesson to be more extensive; for the intervention 'to be a little bit more visual'; and they would like the images to be better organised.

A Year 4 teacher suggested a resource should be provided to start the lesson with an 'oh wow!' effect. He supplemented his own images to use in this way. Teachers also suggested sharing resources among schools.

Content

Teachers felt that the content of the lessons could be made more interesting, and that it would be helpful to establish links with the IPC (International Primary Curriculum) in a structured and consistent way. For example, the children had an 'Aha!' experience when they discovered the connection between the IPC lesson on canals and the lesson on irrigation in Egypt. Teachers would also like to have the opportunity at the beginning of the lesson to explain the bigger picture to contextualize what comes later. For example, one of them explained that in a geography class, she first had to show the map of the UK before she moved on to discuss Wales and Northern Ireland.

What teachers thought of the programme

Teachers were invariably supportive of the programme. Any initial apprehension quickly subsided and teachers adopted the programme enthusiastically. Teachers also reported that there was initial resistance from the pupils, but they gradually grew to like it as they became familiar with the routine.

Some teachers reported that the pupils were very keen and that they liked the activities. For example, children said: 'Can we do the history thing again?' The teacher also reported that pupils were enthusiastic about acquiring knowledge and expressed surprise that pupils liked the traditional format of the intervention and its focus on knowledge acquisition.

Teachers felt that the 'study content was too complicated to begin with, but the pupils gradually grew into the programme as their reading ability and cognitive capacities developed through the year'. They suggested that the texts could be made simpler at the beginning and gradually develop from there.

As both Years 3 and 4 used the same texts, teachers felt that the lack of differentiation between year groups might be why Year 3 pupils found it harder to appreciate the topics. The texts used were deemed more suitable for Year 4. One Year 3 teacher said that although it was great that the students were being given '*some knowledge of things outside Walthamstow*', many students found the concepts hard to access.

Another common comment from teachers was the lack of differentiation. They were concerned about the intervention's lack of 'lower differentiation' for the lower attaining pupils.

Teachers also talked about how the intervention differed from the topic-based, skills-based and child-centred teaching that mostly dominated their perception of good practice. The whole-class teaching approach was a departure from what they had previously been taught, but they believed that there was a place for this. The Year 4 teachers also shared the same sentiments and commented on how they had been surprised by how much their classes enjoyed the structured nature of the intervention.

Some teachers thought the structured curriculum was a good preparation for the linear approach of the new curriculum. Teachers perceived the programme implementation in the context of wider debates about the coming new curriculum that reopened discussions about the right balance between whole-class teaching and individualised, child-centred pedagogies. Thus the program also gave an opportunity for teachers to reflect on their teaching methods and ponder on when and how often whole-class teaching would be an appropriate practice in their teaching.

Teacher survey results

TCC carried out its own teacher survey to assess the impact of the programme. The survey shows that teachers were generally positive about WWR. Teachers reported that pupils were using and understanding a wider range of vocabulary. There was also transference of learning to other subjects. However, teachers did not think that their pupils had made more than expected progress in writing, reading and comprehension. Teachers were also ambivalent about whether the programme increased pupil interest in history and geography.

However, the results of this survey have to be read with caution because of issues with the validity and reliability of the questionnaire items. For example:

- Questions about whether pupils liked history or geography have low validity as they are asking teachers questions that only the pupils themselves can answer.
- Some of the questions were also impossible to answer as they required evidence of measurements which were not carried out.
- The questions about significance are also not reliable because what is significant to one teacher may not be significant to another. Some teachers put a question mark against this question on their response sheet, indicating ambivalence. Teachers' perceptions of significance also vary over time and with individual pupils.
- Questions about evidence of assessment data which show significant growth are leading questions. (Interestingly only 40% of teachers reported that their pupils had better historical/geographical knowledge as a result of the programme than they would without, but 93% said their pupils were aware that they were gaining more knowledge.)

Summary of survey results (n = 15)—responses as a percentage

Question items	Yes	No	Don't know
Do your pupils understand a wider range of vocabulary?	100	-	-
Do your pupils use a wider range of vocabulary?	73	13	13
Have you seen evidence of their use of WWP vocabulary in other subjects?	73	27	-
Have you seen evidence of pupils using WWP vocabulary in their own writing?	73	20	7
Are your pupils aware they are accumulating a wide range of vocabulary?	60	13	27
Are your pupils gaining more knowledge in History and Geography than they would using your school's own curriculum?	40	13	47
Are your pupils aware they are accumulating a great deal of historical/geographical knowledge?	93	-	7
Are your pupils more interested in Geography?	47	7	47
Are your pupils more interested in History?	47	7	47
Has your pupils' confidence in reading aloud improved?	60	13	27
Have your pupils made more than expected progress in comprehension?	40	20	40
Have your pupils made more than expected progress in writing?	20	47	33
Have your pupils made more than expected progress in reading fluency?	13	33	33
Do you have assessment data which shows a significant growth in comprehension skills?	47	27	27
Do you have teacher assessment which shows a significant growth in comprehension skills?	47	33	20
Overall, 53% of teachers indicated that the programme 'has had some significant impact'; 47% reported some impact.			

Pupils' views

There was an obvious contrast in pupils' opinions at the beginning of the intervention and towards the end. During the first evaluation visits, many of the pupils did not have much say: some said they did not enjoy the lessons; some were not aware that they were on a programme, saying that they thought it was just part of their regular lessons. When the evaluator asked them what they thought of their geography and history lessons, the children looked puzzled.

Children appeared to understand the intervention better towards the end. Many said they liked it more at the end as they became familiar with the structure and the routine. On the whole, pupils were positive about the intervention. Many were keen to tell us what they had learnt. Pupils spoke about liking to learn new things. One Year 3 girl talked about a battle between the Spartan navy and the Athenian army and said that when her mum asked her what she had learnt at school she always told her about these lessons. When asked what they had learnt in history, these were some of their comments:

> *'It tells you about the past what weapons they used when they fought and what they used to build their houses.'*

'The lesson helps us to learn new words, to understand what words mean and how to pronounce them.'

The majority of pupils preferred geography to history as it was more hands-on: they could play with the globe and identify places in the atlas. They also liked the pictures and thought there should be more pictures. Pupils also reported learning the pronunciation of new words.

Pupils had mixed opinions about the structured teaching. While they thought the very structured lessons where everyone worked at the same pace was good for team work, it could be frustrating for the better readers as they found the process rather slow. Some talked about being tempted to read ahead and start the exercises before everyone else. Pupils also recognized that it was not necessarily effective for 'some people who have learning difficulties, as they could not always follow the lesson'. The highly structured lesson also did not allow for peer support. Some higher attaining pupils who were used to helping their peers found it unusual that TCC modules did not give space for them to do that.

The pupils suggested having more interesting extension activities for those who finished early to keep them occupied. They would like to see a more varied range of exercises. A number of interesting suggestions were proposed, including greater use of crosswords, wordsearches and puzzles.

As many schools were very much focused on improving literacy skills, the students complained about being overwhelmed with lots of literacy exercises, and they argued that at least the WWR booklet was more exciting than other writing assignments.

Test Administration

Administration of tests varied across schools. Test administration guidance was largely adhered to. In one school, teachers had a briefing beforehand and read through the test administration instructions. Classes had a '*"Do not disturb" sign on the door'* and assemblies for the two year groups were suspended for the day. Testing went smoothly. Children were well-behaved, and kept to the time. Teachers also arranged some quiet activity for children who finished the test before time. Appropriate time was given for each section with relevant breaks in between. There were six forms in this school and all the teachers knew what they were doing.

In another school the teachers were less prepared. Part of the reason was an Ofsted inspection immediately prior to the tests: teachers were busy preparing for this and only had time to read the test guidance on the morning of the test. One of the teachers was not sure whether to take the test seriously or not and whether she was allowed to help the children. Testing rules were flouted as administration instructions were not adhered to. The teacher did not keep to the time and did not arrange for quiet activities for the children. As a result, the children were disruptive and the teacher spent a lot of the test time disciplining pupils.

In most schools, teachers arranged pupils as they would in a normal exam, but in one school, pupils were seated in pairs and there were instances where pupils were seen consulting with each other. In all the schools observed, the test was invigilated by two members of staff (a teaching assistant and a teacher).

There was also a tendency in some schools (both control and treatment) for teachers to give verbal positive encouragements to pupils for correct answers, such as 'Well done!', 'I see some very good writing'.

In a number of schools we observed a few children that were distressed about the test, with some refusing to take it. The heat of the unusually hot summer did not help, and we also noted that children identified with SEN or socio-emotional and behavioural problems were finding it particularly difficult to access the test.

In other schools, pupils with SEN were given a lot of assistance and encouragement. One TA got the pupils to tell them the answers and then wrote the answers for them on a small personal white board. The pupils then copied the answers into their test booklets. As spelling was not penalised in the comprehension section, this was not a major issue. One of the pupils had little English, having just arrived in UK a couple of weeks before. In another school, a TA acted as a scribe for a boy who had his hand in a plaster.[1]

It was also observed that in some schools pupils were constantly seeking assistance. In a treatment school, the TA asked the pupils who were stuck to read the passage to her as a form of encouragement, but did not give prompts or clues.

Teachers' feedback about the GL Assessment test

There were some reservations about the long version of the test: teachers were concerned that the children had only recently completed school assessments, and they felt that the two-hour long PiE test was too strenuous for them. Most schools staggered the tests over two days or more. The long form of the PiE was found to be trying for some pupils. One school decided not to complete the test because it did not think that the children should be put under unnecessary stress, particularly since they did not intend to continue with the programme.

However, teachers were generally positive about the tests: many commented that the writing component was good practice for the children, and felt that it provided a way of levelling, comparing their own assessments with that of the standardised tests; the tests were age-appropriate and not fundamentally different to other tests (such as SATs); the multiple-choice questions were suitable for the lower ability children.

Some teachers, however, raised the concern that the tests might not necessarily assess the skills pupils learnt in the programme, which were very much fact based.

Data collection issues

There were a number of delays relating to the collection of post-tests, pupil data files, and attendance records for a combination of reasons: pupil absences at the end of term meant that test collection was delayed to ensure that all pupils took the test; there were also other end-of-year activities to compete with—some schools had to rearrange their school activities to accommodate the testing. Data collection relating to The Curriculum Centre programme was ignored in some cases. In one school, the staff refused to populate the pupil data.

There were also issues relating to sending completed answer booklets for marking: one school did not arrange for a courier to pick up the tests, and another school mislaid the test booklets. Both issues were resolved.

Implementation

In this section we consider the positive factors that supported the implementation of the programme, and those that were perceived as barriers.

Factors supporting the implementation of the intervention

---

1 Note: Prior to the testing, evaluators conveyed to the programme deliverer that SEN children should attempt the tests, that incomplete tests from SEN pupils were acceptable, and, furthermore, that everything should be done to minimise distress to the children. For these children, and those with limited language ability, evaluators assured schools that TAs or teachers were allowed to read to the children and, if necessary, explain the questions, but that they should not help them with the answers. Special arrangements were also recommended for children with visual impairment: test questions were to be photocopied and enlarged, and answers might be written down with the help of a teacher or TA.

Provision of ready-made teaching resources

The well-developed and structured curriculum, with ready-made teaching resources and teaching aids, ensured consistent and easy implementation. WWR was presented as a complete package with all teaching resources provided and ready to use. Teaching aids like globes, atlases and sample slides/pictures were all part of the intervention. This was attractive to teachers because it meant little planning or thinking about delivery. Activity worksheets were included in the student workbook and ready to use.

Highly sequenced and structured curriculum

The highly prescriptive and sequenced programme meant that there was minimal preparation for teachers—teachers simply followed the procedure for every lesson. The texts had been pre-selected and the guided questions, as well as prompts on how, when and where to ask those questions, were in the teachers' handbook. Children liked the routine of the programme.

Relevance to existing curriculum

WWR fitted in well with the existing literacy curriculum and it was seen as relevant to what teachers were already doing. Teachers did not see the programme as an add-on, but as something that supported their literacy curriculum. Teachers were therefore receptive to the programme.

Teacher quality

Successful implementation also depended very much on the teacher's enthusiasm and effectiveness in making the lessons their own. As one teacher put it:

> *'The success of the lesson depends on the amount of effort the teacher is prepared to put in.'*

Teachers have to be enthusiastic and able to inspire and motivate pupils' interest in the subjects. As the curriculum is very factual, the ability of teachers to engage pupils is crucial in making lessons work.

The more successful lessons tended to be those where teachers related to the pupils' learning and current level of attainment, and adjusted the programme to suit their style of teaching. In contrast, the less successful lessons tended to be those where teachers followed the programme rigidly and did not engage in in-depth discussions of the content material, or add their own personal touches to the presentations. Some appeared to be concerned more with controlling the class. Such lessons were rather less inspiring and pupils did not appear stimulated.

Intensive training of teachers

The training of teachers in the delivery of the programme was crucial to successful implementation. Several teachers mentioned that the demonstration lessons given by the programme developer were very helpful.

However, it was felt that the training should include not only the theory of learning and delivery, but also content knowledge. Successful lessons were those where teachers brought in their own understanding and experiences relating to the topics. In many instances, observations suggested that the curriculum was not taught very well because teachers were deficient in general knowledge: providing teachers with various information sources relating to topics covered in the programme would address this issue. Teachers welcomed the opportunity to share resources and ideas.

Barriers to the implementation of the intervention

Lack of differentiation between year groups

Schools reported that there was an obvious maturity gap between the 7–8 and the 8–9 year-olds. The topics and level of activities at its current configuration seemed to suit the Year 4 age group better. The older children were better able to understand and grasp the knowledge and enjoyed the topics more. Correspondingly, Year 3 pupils did not seem to get much benefit from the programme. Teachers suggested that choosing topics appropriate for the younger children might help.

Lack of differentiation between higher and lower attaining pupils apart from the extension activity

The lack of differentiation in the materials used in the text was a common criticism. Some pupils, particularly the younger ones and EAL children, found the topics challenging and inappropriate for their ability level. Those on levels 2b and 2c, for example, struggled with the lessons. Teachers suggested simpler questions and a different range of keyword vocabulary for the weaker ones. In the lessons observed, the higher ability pupils completed the writing activities with ease while the lower ability pupils appeared to be struggling: they were either not writing at all, or were only answering one or two questions. This was particularly true of Year 3 history classes. Teachers also expressed concern about provision for SEN and EAL children (currently, the programme does not cater for those children). One teacher remarked that some of her Romanian EAL pupils could not be included in the lessons at all.

In a number of lessons observed, it was clear that the less able pupils were easily distracted, and many simply abandoned the writing activity because it proved too challenging for them. They started chatting or doing something else.

Teacher's level of general knowledge

Teachers' lack of background knowledge was perhaps the biggest hindrance to the successful implementation of the programme. There were a number of instances where teachers made factual errors because of their own lack of general knowledge. For example, in a geography lesson one teacher caused some confusion when he told the children that two pictures on the screen both showed sand dunes, when it was quite clear that one of them was a rocky coastline. In another geography lesson, pupils were confused about tides and waves as they were all described as 'movements of water up and down'.

Lack of images

Insufficient slides, or pictures that teachers could use to support their lessons, was another limiting factor.

The rigid structure of the curriculum

Some teachers found it hard to adapt the rigid structure of the curriculum to their teaching style, with lessons appearing forced and contrived: in such cases there was little attempt to bring forward the discussion or to develop lesson themes further.

Alignment with the existing curriculum

One challenge highlighted by some teachers was the need to tie the programme in with the existing curriculum. Teachers mentioned that 'squeezing it into an already tight curriculum' was a problem.

Lack of autonomy in the classroom

The structured nature of the programme meant that there was little room for innovation and autonomy in the class. Some teachers found it a struggle to keep very close to the text, and it was observed that some teachers had difficulty handling quite thought-provoking questions from children. In some cases

it appeared that teachers thought that they must always have the answers to everything, rather than making it a learning experience.

High teacher turnover

The high staff turnover in a few schools also made it difficult for some to implement the programme fully. In one school, almost all the teachers who were originally trained—including the lead person—left (Table 16). In these cases the programme developer conducted training for the new teachers, however in our final visits we found untrained teachers who had received unofficial training from the school lead. There were minor departures from the TCC protocol in the lessons conducted by these teachers: one teacher, for example, appointed pupils to read the text to the class instead of reading the text himself; another thought the keyword exercises, which were crucial in building vocabulary, were optional and so the children missed most of these exercises.

Table 16: Turnover in relevant staff in the treatment schools

School	Number trained	Number leaving
School 1	2	2 plus headteacher
School 2	5	4
School 3	12	6
School 4	4	1
School 5	5	1
School 6	6	1
School 7	-	No change
School 8	-	No change
School 9	-	No change

Note: the developers have not provided the number of teachers in schools 7, 8, and 9.

Is the intervention attractive to stakeholders?

As noted, the complete package approach (with lesson outlines and resources) was attractive to teachers as it reduced their workload. It also complemented other activities in the curriculum. Teachers also reported evidence of impact on some aspects of pupils' learning. It trained pupils in the discipline of answering questions in full sentences, and in comprehension skills.

The curriculum included content knowledge about geography and history. Pupils liked the routine and the structure of the lessons. Many explained that they enjoyed the lessons because they were learning new knowledge—something that they would not have learnt otherwise. Pupils talked about the excitement of learning about the planets, oceans and ancient civilisations, however not all pupils shared the same views: some pupils found the curriculum 'boring' and challenging.

Five of the treatment schools have confirmed that they are continuing with the programme. Two further schools not involved in the trial have also indicated that they would like to adopt the WWR curriculum, one of them intending to use the geography curriculum with their Year 5 pupils.

Outcomes

Changes in teachers' behaviour

There were apparent improvements in teachers' behaviour, confidence and teaching styles between the first and the final observations: lessons tended to be smoother and teachers were also more confident, with some developing their own style of delivery. One teacher, for example, introduced themes and discussions that anticipated the keywords and the extension questions. This meant that when it came to the open questions, very interesting and complex discussions broke out which he brought together at the end of the lesson. The writing exercise became not just about comprehension, but also an exercise in thinking about the concepts and issues relating to the topic of the day.

Impact on pupils' learning

Some teachers reported that the programme had a perceptible impact on children's composition skills. Children were learning to use full sentence answers which the teachers thought was particularly useful in tests. Pupils learnt to use sentence starters, and this had been applied across the board. The impact had been seen in other aspects of the curriculum, for example, using the vocabulary learnt in other writing. Teachers also said that the practice of answering in full sentences had filtered in to other work, and the benefit of the intervention could be seen in students' SATs papers.

Teachers also spoke about how students had become more confident in using technical terms. They agreed that in Year 4 it had been possible to unfold more conceptual discussions around the keywords than in Year 3.

A deputy head spoke about how pupils often excitedly told her what they had been learning in the lessons. She was certain that the intervention had 'contributed significantly'. But the school also had a commitment to daily guided reading and parental support interventions. Because of all these other interventions it was difficult to isolate the impact of this intervention. As one Year 3 teacher said: 'It is only a little piece of the puzzle'.

There was also evidence that pupils were thinking about the vocabulary they learnt during the lessons. For example, pupils asked one of the evaluators whether it was proper to use 'death' or 'died', such as 'after his death' or 'after he died'. Another teacher reported that pupils also attempted to use keywords learnt in their writing. For example, words associated with the compass points (N, S, E, W), seasons, towns and cities, farming, invention and environment.

Further evidence of learning could be seen in pupils' workbook exercises. For example, pupils demonstrated their grasp of quite abstract concepts such as 'democracy' and 'government' in the sentences they made. One child wrote about how they had had difficulty deciding what to do, so they had had a vote and made the decision democratically, illustrating the pupil's understanding of the word, 'democratic'. However, such concepts proved challenging for Year 3 pupils. The comprehension exercises also trained pupils to look at the text for answers. A Year 4 pupil added that the lessons gave her interesting things to write about, and she enjoyed the space the keyword and extension questions gave her to develop her own ideas:

> *'You can go back to the text and read, not just write with your head down…*
> *You can write your own things. It makes us think a lot. There is a lot of*
> *"What do you think?". It's our opinions that we write.'*

Another said of the class in general that she liked how they 'get to discuss'. Children were reportedly gaining more knowledge about the world around them. Some teachers felt that the programme had helped hone pupils' comprehension skills and trained them to assess a piece of writing.

Impact on other aspects of learning was also noted. Some teachers noticed that pupils were listening better and working a lot faster. One teacher noticed, for example, that when a new pupil arrived there was an obvious difference between the child and the other pupils, but that once exposed to the programme the child began to catch up. Pupils reported that what they had learnt in WWR had helped in other lessons, particularly IPC. Some topics overlapped, for example, lessons on 'volcanoes'.

Pupils' responses were mixed regarding the usefulness of answering in full sentences. One said that it did not help in literacy as 'literacy is different' and 'there are no questions like this'; another said it had helped with 'big writing'; and one Year 4 pupil said that writing in full sentences was good practice as it had helped them with comprehension tests where marks might be lost for not answering in full sentences.

What the WWR pilot emphasised was pace. One teacher described it as: 'Let's go! Boom boom boom!'—it encouraged 'pace', 'whipping it out', she said. This was more apparent in some lessons than in others.

A general observation was that the vocabulary activity was not able to show pupils' understanding of the concepts—many pupils merely repeated the question when asked to make sentences using the word.

Area of concern

In some cases, teachers did not appear to have sufficient information or subject knowledge to discuss material, even to a very limited extent, resulting in the danger that factually wrong information was being transmitted to children. This was observed in a number of lessons.

Fidelity

The structure and protocol of the Word and World Reading curriculum as demonstrated in the training was generally adhered to. There were slight departures from the programme observed with two newly hired teachers who did not receive the official training from the programme developer. One teacher asked pupils to read the passage instead of reading it himself. Another teacher mistakenly thought the vocabulary exercises were optional, and pupils in that class missed those activities which were crucial to the programme. Generally, teachers followed the sequence, the prompts, and the on-the-spot marking as suggested by the curriculum. However, there were instances where pupils' books were not marked, or where mistakes were not spotted and corrected. There were variations among teachers in how closely they marked pupils' books. In most cases mistakes were circled and pupils corrected them. In a few instances some obvious grammatical and spelling mistakes were not spotted.

The pre-test was not observed by the evaluators, and the papers were marked by GL, the independent test developer.

The post-test was taken seriously in the majority of schools, including the control schools. Pupils took the test under strict exam conditions. As this was a two-hour test, many schools staggered it over a number of days, and in short sessions as recommended by the test developer. One treatment school did not administer the full test because they did not intend to continue with the programme and thought they should not put the pupils through the test.

EAL children (particularly those who had recently arrived in the country), SEN children, and those with social-emotional and behavioural problems found the test conditions particularly trying: all were encouraged to take the test and every effort was taken to make it easy for them. Schools were assured that there was no pressure for such children to complete the test, but that every child should be given the opportunity to attempt it. According to the test developer, a certain degree of assistance was allowed for these children: TAs were, for example, permitted to read a question or a passage to them. In a few cases we observed for the extended comprehension section, the children verbalised their answers to the TA who then wrote them down on a tablet, and the children then copied the answers into their booklet. This was deemed acceptable as the extended comprehension questions were not marked for spelling and grammar. Where children had assistance from TAs, a note was made to that effect. Teachers and TAs tried their utmost to assist EAL and SEN children with the tests. Some children with behavioural difficulties were taken out of the classroom to lessen the anxiety for them.

The marking of the test scripts, however, suggested a number of irregularities, mainly in the extended questions. There was evidence of coaching in some schools where pupils had two sets of answers—

one on draft pieces of paper, and one in their test booklet. In some classes, pupils' answers were very similar.

Most of the schools completed the multiple-choice sections (grammar, spelling and comprehension), but three did not fully complete the extended reading comprehension and writing composition.

Formative findings

Teacher support

Teachers at different levels may require different levels of support. An experienced teacher may need minimal support, but an average teacher may need more specific guidance. Modelling of lessons is useful. Teachers found the demonstration lessons by the programme developer very helpful.

Greater collaboration among teachers

There were apparently some good lessons, but these were rarely shared. More opportunities could be created for teachers to observe each other's lessons, share experiences, resources and tips about how to make the lessons interesting. One recommendation is for teachers to attend workshops to share and produce teaching materials. Teachers relished the idea of sharing ideas and resources.

More intensive teacher training

Intensive training of teachers is necessary for successful implementation. It is likely that training would be more effective if it included sections on subject and content knowledge, with perhaps a couple of sessions on the geography and history topics covered in the curriculum, in addition to information about the theories and principles behind the programme. It may be useful to direct teachers to relevant and useful websites—sources that teachers can dip in for quick access to information. It would also be beneficial to have more suggested ideas on how teachers can make the lessons more interesting. In lessons with the weakest implementation, the programme could appear rather contrived and lack variety.

Greater differentiation

The curriculum as it is currently presented does not differentiate children of different ability or year group. EAL children and some children identified with special educational needs found it difficult to access the curriculum. Perhaps there could be differentiation in terms of topics and range of vocabulary, or in the selection of text passages that could be tailored more appropriately for children of different ability and year group. It was observed that some of the concepts were way beyond the children's ability to grasp.

More varied workbook activities

Teachers and children had asked for a wider range of activities in the workbook, such as more puzzles, crosswords and word searches. These workbook activities could be differentiated to cater for children of different ability. It was observed that while some children struggled to do the exercises, others found them too easy and boring.

More attractive presentation

Since the programme was specifically for children in primary school, pupil textbooks could be presented more attractively. In our interviews, children consistently asked for more coloured images in their textbooks. The front cover page could be made more attractive to encourage interest in the subjects.

Greater autonomy

Although the highly structured and prescriptive programme was well received by teachers, it could stifle creativity and innovation. The programme as it is currently packaged offers little opportunity for teachers to be creative. Only a handful of teachers attempted to be creative by supplementing the lessons with their own video clips and slides.

More scope for in-depth discussion of topics

According to Hirsch, teaching facts to pupils gives them the foundation to think critically and gives them opportunities to apply the knowledge and to question the facts. Hirsch's original idea for the sequential curriculum was a curriculum that covers 50% of content and 50% for wider discussions of the topics covered. The curriculum should be designed to allow for such discussions (Core Knowledge, 2014). One recommendation for the WWR curriculum is perhaps to have fewer topics but more in-depth coverage, with topics developing as the child moves up each level, progressing from simple concepts to more complex ideas.

It may also be helpful for the first lessons to begin with a discussion about what 'geography' or 'history' is about. During our school visits, children asked about ancient civilisations and how we knew what Mesopotamia was like indicating that an introductory lesson would be useful. Google Maps and satellite images could also be used to teach geography.

Tests

The post-design decision to use the extended version of the test (to assess writing ability) for the post-test was not well received by schools. It was perhaps too long for 7–9 year olds, and many schools felt that this was encroaching on their teaching time. Consequently, a number of schools did not attempt all sections of the test. Schools could be better prepared for such extended assessments and primed beforehand.

Training relating to evaluation

Future evaluations could consider a pre-intervention workshop for schools about the process of evaluation: the importance of complete data, keeping attendance records (for assessing impact of dosage), and commitment to testing. The implications of failure to comply for the security of the evaluation findings could be explained. In this evaluation, the evaluators requested simple attendance records from schools (via the developer) but too few were received to be useful for analysis.

Control group activity

Pupils in the control schools continued regular lessons as normal. They were to receive the intervention a year later, but to avoid post-allocation demoralisation they were offered £500 to cover for teachers while they received training to use the resources. The evaluators have no evidence that these schools adopted alternative literacy intervention once the allocation had been revealed (although this is always a possibility with school-level allocation).

Impact Evaluation

Aims

The impact evaluation linked to this pilot study was primarily aimed at providing an estimate of the likely magnitude of effect of Word and World Reading on reading comprehension of Year 3 and 4 pupils in primary schools with a high proportion of disadvantaged children. This estimate can be used both to inform a decision regarding whether there is sufficient promise for a larger trial, and if so, to help assess how large that trial would have to be in order to stand a reasonable chance of detecting the impact. The results aimed to be indicative only.

Timeline

June–July 2013
The pre-test was conducted in 17 of the schools that had signed up for the programme.

August 2013
Training of staff began.

September 2013–June 2014
Delivery of WWR began in treatment schools. Evaluators conducted visits to treatment schools throughout the intervention to observe the delivery of the lessons.

July 2014
The conduct of the post-test continued throughout the month. Most schools completed the test in the first two weeks of July.

August–October 2014
Marking of post-test long answer scripts by evaluators. Chasing up and collection of school leavers' test scripts.

November to December 2014
Matching of pre- and post-test data, analyses of outcome measures, and writing up of evaluation report.
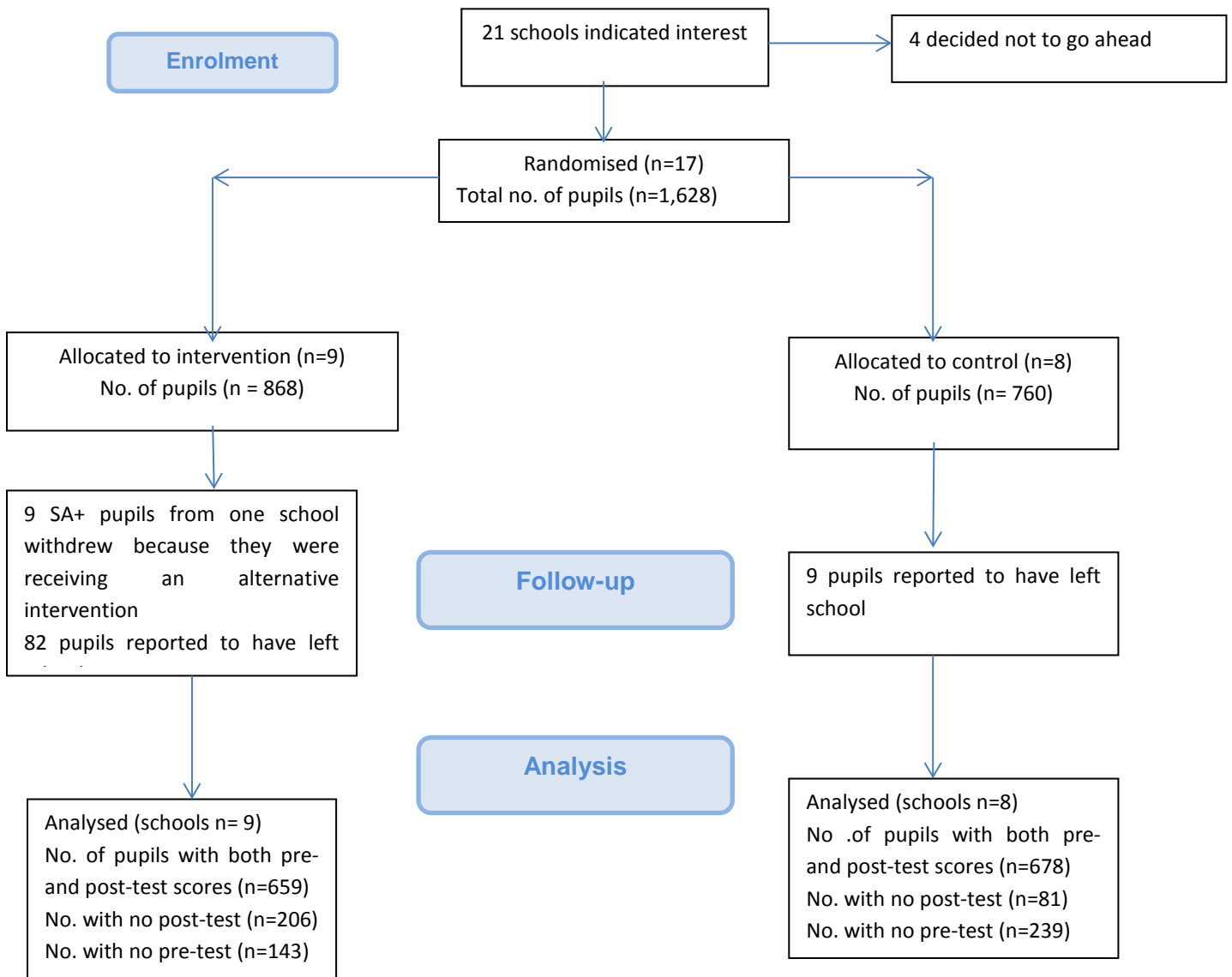
Participants

In the first call, 21 schools registered interest in the programme and signed up. However, when the programme started in earnest, four schools decided not to go ahead because they did not feel that they could commit to the programme. In the event, the project team successfully recruited 17 schools: of these, 9 were randomised to receive the intervention immediately and the other 8 schools, forming the control (waitlist), were to receive the intervention a year later.

In total, 1,628 pupils were signed up. Nine pupils from one treatment school withdrew subsequently because they were being dealt with under School Action Plus (SA+) and were receiving another intervention deemed more appropriate for them. One treatment school withdrew at the beginning of the programme after pre-test because they had a new head who did not think that the WWR programme aligned with their new curriculum. This school provided post-test data, and is included in the intention-to-treat.

One school did not complete the pre-test as they did not have the time to do so, but continued with the intervention with the agreement of the developers and the EEF.

The flow chart below tracks the number of schools/pupils from initial enrolment and allocation to analysis. In the final analysis, 1,340 pupils were analysed from 17 schools.

Participant flow diagram

Enrolment

21 schools indicated interest → 4 decided not to go ahead

Randomised (n=17)
Total no. of pupils (n=1,628)

Allocated to intervention (n=9)
No. of pupils (n = 868)

Allocated to control (n=8)
No. of pupils (n= 760)

9 SA+ pupils from one school withdrew because they were receiving an alternative intervention
82 pupils reported to have left

Follow-up

9 pupils reported to have left school

Analysis

Analysed (schools n= 9)
No. of pupils with both pre- and post-test scores (n=659)
No. with no post-test (n=206)
No. with no pre-test (n=143)

Analysed (schools n=8)
No .of pupils with both pre- and post-test scores (n=678)
No. with no post-test (n=81)
No. with no pre-test (n=239)

School characteristics

The schools allocated to treatment and control groups were diverse and reasonably well-balanced in terms of type, intake and results (Tables 1 and 2).

Table 1: Summary characteristics of treatment schools (2013)

Type of school	Age range	Enrolment	%SEN	%EAL	%FSM	% ethnic minority	KS2 average point score	Ofsted effectiveness (date of inspection)
Academy Converter	3–11	200	11	58	41	83	28	Good (2010)
Academy Sponsor	4–11	450	6	30	54	44	23	Inadequate (2014)
Community	7–11	350	12	60	25	83	28	Not available
Academy Sponsor	3–11	300	20	84	46	96	26	Not available
Academy Converter	3–11	1050	3.8	39	43	97	31	Good (2013)
Academy Sponsor	3–11	200	17	9	42	28	28	Requires improvement (2014)
Academy Sponsor	3–11	400	10	26	52	30	26.	Good (2014)
Academy Converter	3–11	450	14	78	40	95	29	Outstanding (2013)
Academy Sponsor	4–11	200	13	25	48	45	26	Not available

Note: figures have been rounded.

Table 2: Summary characteristics of control schools (2013)

Type of school	Age range	Enrolment	%SEN	%EAL	%FSM	% ethnic minority	KS2 average point score	Ofsted effectiveness
Academy Sponsor	3–11	150	14	8	69	35	24	Inadequate (2014)
Academy Converter	4–11	200	5	12	27	29	30	Outstanding (2009)
Academy Converter	5–11	780	14	60	31	85	29	Outstanding (2006)
Academy Converter	3–11	450	5	4	35	6	27	Good (2012)
Academy Converter	4–11	250	11	5	28	10	26	Good (2010)
Academy Sponsor	4–11	200	13	12.7	24	22	28	Good (2014)

| Academy Sponsor | 3–11 | 900 | 9.1 | 32 | 52 | 93 | 27 | Satisfactory (2014) |
| Academy Sponsor | 4–11 | 350 | 18.5 | 35 | 48 | 23 | 26 | Satisfactory (2014) |

Note: figures have been rounded.

Pupil characteristics

Perhaps because of the small number of cases randomised (17 schools), the initial balance between pupils in the two groups is poor (Table 3). The treatment group has markedly more FSM-eligible and ethnic 'minority' pupils and those speaking a first language other than English. The control group has markedly more pupils listed as having a special educational need. These kinds of differences could influence the baseline pre-test results, which is one of the reasons why the 'gain' scores in PiE are used to assess the results of the trial.

Table 3: Characteristics of pupils in treatment and control groups at the outset

Pupil characteristic	Treatment	Control
Male	322 (49%)	350 (51%)
FSM-eligible	237 (36%)	183 (27%)
SEN	112 (17%)	171 (25%)
EAL	298 (45%)	164 (24%)
Non-White British	505 (77%)	308 (45%)

The two groups were reasonably well-balanced in terms of the pre-test scores (despite the concerns over their background characteristics noted above). They were also reasonably well-balanced in terms of the post-test scores, and so also in terms of gain scores.

Attrition

A total of 287 post-test results were missing (206 from the treatment schools, and 81 from control schools). Most of the missing cases came from one treatment school which accounted for 43% of the total missing cases (17.7% overall). There is also an imbalance in the kind of pupils who failed to take the post-test. For example, there were proportionately more boys and SEN pupils who missed the post-test from control schools. One possible reason could be that schools routinely excluded SEN pupils from tests.

In line with the intention-to-treat principle, we attempted to track all pupils who left school so that post-test data could be collected from them. Schools provided the names of such pupils, and these were tracked. Names were provided from both treatment (n = 81) and control schools (n = 9). We could only speculate the reasons for the disproportionate information about dropout. One possible reason could be that control schools did not think it was necessary for them to inform us of dropouts as they were not actively involved in the intervention. The programme developers may not think it mattered too much to insist that schools furnish this information. It is possible that the pupil population in the two groups was very different. There were proportionately more ethnic minority and EAL children in the treatment schools (see Table 3). Some of these children may have been in a school temporarily while they sought longer term accommodation.

Where pupils could be traced, emails were sent to the destination schools to request their assistance in administering the test. The purpose of the project and the importance of full and complete data were explained. Following these emails, test booklets with explanatory and administration guidance notes were sent out. These were followed up with phone calls. Despite our best efforts, a number of schools indicated that they would not be able to help out because of the length of the test (two hours).

A common reason given was that it was not fair to take the pupils out of lessons for two hours, and also most schools had just had their SATs. Schools did not feel that they should put pupils under any more stress than was necessary. This is an indication that the long form of PiE insisted on by the developers after the trial was underway should not be used in future.

Many of those who left could not be traced either because they had left the country, were home-schooled, or for confidential reasons their whereabouts could not be revealed. In fact, many more children left school than were reported, and more arrived and often sat the post-test with their peers for convenience.

Outcomes and analysis

Using the evidence available, the intervention has had no discernible impact on PiE progress (Table 4). In summary, the treatment group was slightly ahead at the start and slightly behind at the end.

Table 4: Raw score results, gain scores, all schools

	N	PiE pre-test	Standard deviation	PiE post-test	Standard deviation	Gain score	Standard deviation	'Effect' size
TCC	659	22.9	8.4	47.5	15.7	24.7	12.6	-0.03
Control	678	22.2	8.9	47.8	16.3	25.1	12.9	-
Total	1,337	22.6	8.6	47.7	16.0	24.9	12.7	-

Note: One control school provided only post-test scores so the gain scores have N of 565 in the control group.

The mean raw score for the school with no pre-test scores was 50.6 (SD 14.2) in the post-test, somewhat higher than the overall average (Table 4). This means that using the post-test scores for this school may be reducing the apparent effect size of the intervention. This is assessed in Table 5 which excluded this school from analysis. The overall finding is the same—no evidence of discernible beneficial impact. The average score for the post-test in the control group only drops from 47.8 (Table 4) to 47.3 (Table 5) so the remainder of the analysis uses all available scores.

Table 5: Raw score results, all pre-test schools

	N	PiE pre-test	Standard deviation	PiE post-test	Standard deviation	Gain score	Standard deviation	'Effect' size
TCC	659	22.9	8.4	47.5	15.7	24.7	12.6	-0.03
Control	565	22.2	8.9	47.3	16.6	25.1	12.9	-
Total	1,337	22.6	8.6	47.4	16.1	24.9	12.7	-

Exactly the same result emerges if standardised age scores are used for pre- and post-test results, instead of raw scores (Table 6).

Table 6: Standardised age score results, all schools

	N	PiE pre-test	Standard deviation	PiE post-test	Standard deviation	Gain score	Standard deviation	'Effect' size
TCC	659	99.5	15.2	93.4	14.5	-3.1	13.0	-0.02
Control	678	99.3	15.8	96.8	15.7	-2.9	13.1	-
Total	1,337	99.5	15.5	96.6	15.1	-3.0	13.0	-

The results are slightly more positive for FSM pupils considered in isolation (Table 7) but, as with other findings, this finding is not secure as the number of pupils is small, and FSM pupils were not randomised as such.

Table 7: Raw score results, all schools, FSM pupils only

	N	PiE pre-test	Standard deviation	PiE post-test	Standard deviation	Gain score	Standard deviation	'Effect' size
TCC	237	22.9	8.6	46.2	15.9	23.3	12.5	+0.06
Control	157	20.4	9.0	42.8	16.3	22.5	14.2	-
Total	394	21.9	8.9	44.7	16.1	23.0	13.2	-

An additional analysis by sex shows that boys did noticeably worse, on average, than girls (Table 8).

Table 8: Raw score results, all schools, boys only

	N	PiE pre-test	Standard deviation	PiE post-test	Standard deviation	Gain score	Standard deviation	'Effect' size
TCC	322	22.1	8.4	45.4	15.8	23.4	12.6	-0.13
Control	287	20.6	9.3	45.7	15.5	25.0	12.2	-
Total	609	21.4	8.8	45.6	15.6	24.1	12.5	-

Note: The equivalent effect size for girls was +0.05.

The results are effectively the same for both age groups in the trial (Tables 9 and 10).

Table 9: Raw score results, all schools, Year 3 only

	N	PiE pre-test	Standard deviation	PiE post-test	Standard deviation	Gain score	Standard deviation	'Effect' size
TCC	304	21.6	8.5	46.4	14.5	24.8	11.2	-0.03
Control	307	21.6	8.7	46.9	15.7	25.2	12.5	-
Total	611	21.6	8.6	46.7	15.2	25.0	11.8	-

Table 10: Raw score results, all schools, Year 4 only

	N	PiE pre-test	Standard deviation	PiE post-test	Standard deviation	Gain score	Standard deviation	'Effect' size
TCC	355	24.0	8.1	48.5	16.5	24.5	13.7	-0.04
Control	258	22.9	9.0	48.9	16.9	25.0	13.4	-
Total	613	23.5	8.5	48.7	16.7	24.7	13.6	-

Table 11 represents the R values for the two regression models, each based on two steps. In Step 1, the pupil background and pre-test scores are included, and then in Step 2 the binary variable for being in the treatment or control group is added. The model is better at explaining variation in the post-test outcome than the gain score outcome, but for both models the bulk of the variation that is 'explained' by the variables in the model is explained at Step 1. Once pupil background and prior attainment are accounted for, very little difference is explained by knowing whether a pupil was in the treatment group or not. This model is not, in itself, any test of causation, but it does act as some confirmation of the headline finding of no impact, even when differences in pupil characteristics between the initial groups are taken into account.

Table 11: Variation explained (R) in two-stage regression model, using two possible outcomes

	Gain score outcome	Post-test outcome
Step 1: background and prior attainment	0.27	0.65
Step 2: intervention	0.28	0.65

For completeness, Table 12 presents the coefficients for all explanatory variables retained in either model. The largest of these by some way are the precise age in months when taking the test, and the pre-test score.

Table 12: Standardised coefficients for the regression model in Table 11

	Gain score outcome	Post-test outcome
FSM	+0.01	-0.09
Sex (female)	+0.08	-0.05
SEN	+0.06	-0.06
EAL	+0.01	-0.10
Ethnicity (White UK)	-0.06	-0.15
Age at pre-test	-1.04	-0.30
Age at post-test	+0.98	+0.30
PiE (pre-test)	-0.56	+0.06
Step 2: Treatment (or not)	-0.03	+0.07

Cost

The costs for delivering the programme (which include staff training and resources) were provided by the programme deliverer. Working on the assumption of 25 pupils in each class, the cost per pupil would be £50 (Table 13).

Table 13: Cost per item for a class of 25

Item	Cost
Globes	£237.65
Atlases	£208.50
Pupil workbooks	£171.60
Teacher workbooks	£7.00
Teacher cover (supply)	£200
Fees for teacher training	£200
Admin costs (estimate)	£200
TOTAL	£1,224.75

Bearing in mind that the teaching resources (but not the pupil workbooks) can be re-used, costs should go down after the initial investments have been made. Once teachers are trained, they need not be retrained. Training is only required where there is staff turnover. No top-up training is required.

Conclusion

Key conclusions
1. The Word and World Reading programme was introduced as intended, and well-received by the majority of primary schools participating in the project.
2. Some teachers felt that the programme had a positive impact on pupil learning, including improved vocabulary and writing skills.
3. In some lessons, teachers' subject knowledge did not appear to be sufficient to support an in-depth discussion with pupils about some of the topics within the programme curriculum. This suggests that additional training or support materials may have been beneficial.
4. The programme appeared to be more successful for older, higher attaining students and less successful for Year 3 students or low attaining students. Greater differentiation, for example adapted vocabulary lists, may have made it easier for lower attaining students to engage with the programme.
5. The study did not seek to assess impact on attainment in a robust way, however the attainment data which was collected did not indicate a large positive effect. This suggests that any future trial of the programme should involve a large number of schools in order to provide a precise assessment of the cost-effectiveness of the programme. It may also be valuable to test the approach over a longer period of time.

Limitations

This pilot evaluation was conducted as planned, and enabled an assessment of the feasibility and promise of the programme. Data was collected through lesson observations and interviews with pupils and staff by the independent evaluators, with input from developers. In total 12 visits were made to 8 participating schools and 56 classes were observed. In addition, eight visits were made to schools in both the intervention and comparison groups when pupils sat the end of project test.

The impact evaluation provided an estimate of the likely magnitude of effect of the Word and World Reading pilot on reading comprehension in order to inform a future trial. The trial was not large enough to provide a secure estimate of impact, but did indicate that any future trial should involve a large number of schools.

One limitation of the impact component was that most evaluations of the Core Knowledge curriculum (which the WWR pilot is modelled on) followed children for three years rather than one year. It is likely that the effects of teaching facts are not immediately reflected in literacy performance: such knowledge builds up over time and remains latent. It would be interesting to follow these children to secondary schools to see if there is a difference between children taught TCC curriculum (which includes geography and history) and those who were not.

A number of schools that implemented the programme did not administer the full test. This could have affected the results. Some schools were under the impression that because they did not intend to continue with the programme, the post-test was not necessary.

The schools selected represent a diverse range of provision. Randomisation at the school level reduced the possibility of diffusion as control and treatment schools were kept separate.

Pupils, especially those from traveller and migrant communities, were also very transient. Movements of pupils in and out of the larger schools tended to be very fluid, and were not always reported to evaluators. In a number of the larger schools as many as a 100 of the original children left during the study. Some had left and returned to their home country, some moved location and could not be

traced, and others were home-schooled. In cases where pupils had left school and evaluators were informed, efforts were made to track them and make arrangements for them to take the post-test in the new school. Some of these schools were very helpful in agreeing to administer the post-test, but unfortunately, in a large number of cases, it was not possible to obtain test data.

There were also schools where almost all the teachers initially trained to deliver the programme left, and a new set of teachers arrived. In some schools even the head changed making it difficult to continue with the programme.

These difficulties inevitably meant that the pre-test group differed significantly from the post-test group.

Interpretation

This evaluation demonstrates that a prescriptive and highly structured programme can be integrated into the school curriculum with minimal adjustments to time-tabling—if perceived as complementary and relevant to the existing curriculum, it can be readily adopted.

However, a degree of flexibility is necessary for successful implementation. To achieve this, intensive training of staff is crucial, not only relating to programme delivery, but also regarding content knowledge and pedagogical skills (since teaching literacy is quite different to teaching subject content). Having a ready-made curriculum alone, as here, is not enough: teachers have to be passionate about the subjects, willing to engage pupils in discussions, and explore beyond what is in the text. Many teachers would need to be shown how this can be achieved. Strong administrative support from the school's management team is also essential for optimal implementation. It was noticed that schools where such support was apparent were more committed to the programme. One school pulled out of the programme due to lack of administrative support as the new head did not 'buy in' to the philosophy of the curriculum.

According to the original idea of this programme, children's learning of factual knowledge needs to be built up over time, so the one-year implementation may not be long enough for the effects to be apparent. Also children need to start early. The strongest evaluations of the Core Knowledge curriculum have shown that positive effects were found among younger pre-school children; results for older children (grades one to five) were not conclusive and less convincing. The intervention relies on teachers imparting knowledge to children, the success of which depends on the ability of teachers to conduct in-depth discussion of the content material with children. Not all teachers in this pilot scheme appeared to be able to do this, and this is perhaps one of the biggest challenges for the intervention.

As a test of feasibility, this pilot demonstrates that it is feasible to conduct a trial involving a large number of diverse schools across geographical areas. The programme was well-conducted and closely monitored throughout.

It was challenging to track participating pupils and teachers in and out of the schools, which may be a barrier for a future study.

Overall, the Word and World Reading curriculum was well structured and schools adhered to the prescribed number of sessions per week. The implementation of the programme was closely monitored by the curriculum developer. This ensured that teachers implemented the programme as they were trained.

Future research and publications

If the programme of research in this area is to continue, the following issues could be addressed:

- What is the optimum number of sessions?
- Would a three-year (rather than a one-year) programme enhance impact?
- Would more in-depth coverage of content, with scope for greater discussions, be effective?
- Are any positive effects resulting from the programme sustained over time?
- What is the best choice of test to reflect the acquisition of content knowledge?

The evaluators plan to produce a peer-reviewed journal article, or similar, based on this pilot trial.

References

Berk, R. and Freedman, D. (2001) 'Statistical assumptions as empirical commitments', http://www.stat.berkeley.edu/~census/berk2.pdf, accessed 3 July 2014.

Carver, R. (1978) 'The case against statistical significance testing', *Harvard Educational Review*, 48, 378–399.

Core Knowledge Foundation (2000) 'In Oklahoma City, a rigorous scientific study shows the positive equity effects of Core Knowledge', http://coreknowledge.org/CK/about/research/oklahoma_study_2000.htm, accessed 24 April 2009.

Core Knowledge (2014) 'E. D. Hirsh and Cultural Literacy', http://www.coreknowledge.org.uk/culturalliteracyck.php

Dartnow , A., Borman, G. and Springfield, S. (2000) 'School reform through a highly specified curriculum: implementation and effects of the Core Knowledge Sequence', *The Elementary School Journal*, 101, 2, 167–192.

Falk, R. and Greenbaum, C. (1995) 'Significance tests die hard: the amazing persistence of a probabilistic misconception', *Theory and Psychology*, 5, 75–98.

Gibb, N, (2014) 'The importance of teaching'. Speech given at ResearchEd conference in London, https://www.gov.uk/government/speeches/the-importance-of-teaching, accessed February 2015.

Gorard, S. (2013) *Research Design: Robust approaches for the social sciences*, London: SAGE.

Hirsch, E. (1987) *Cultural literacy: what every American needs to know*, New York: Vintage.

Iver, M., Stringfield, S. and McHugh, B. (2000) 'Core Knowledge Curriculum: Five-year analysis of implementation and effects in five Maryland Schools', Report No. 50, Baltimore: John Hopkins University.

Lipsey, M., Puzio, K., Yun, C., Hebert, M., Steinka-Fry, K., Cole, M., Roberts, M., Anthony, K. and Busick, M. (2012) 'Translating the statistical representation of the effects of education interventions into more readily interpretable forms', Washington DC: Institute of Education Sciences.

Stringfield, S., Dartnow, A., Borman, G.. and Rachuba, L. (2000) 'National evaluation of Core Knowledge Sequence implementation Final Report', Report No. 49, Baltimore: CRESPA, John Hopkins University.

The NYC Core Knowledge Early Literacy Pilot (2012) http://www.coreknowledge.org/mimik/mimik_uploads/documents/712/CK%20Early%20Literacy%20Pilot%203%2012%2012.pdf, accessed July 2014.

Watts, D. (1991) 'Why is introductory statistics difficult to learn?', *The American Statistician*, 45, 4, 290–291.

Whitehurst-Hall, J. (1999) 'The impact of the Core Knowledge Curriculum on the achievement of seventh and eighth grade students', PhD dissertation, Faculty of Education, University of Georgia, Athens.

You may re-use this document/publication (not including logos) free of charge in any format or medium, under the terms of the Open Government Licence v2.0.

To view this licence, visit **www.nationalarchives.gov.uk/doc/open-government-licence/version/2** or email: **psi@nationalarchives.gsi.gov.uk**

Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned. The views expressed in this report are the authors' and do not necessarily reflect those of the Department for Education.

This document is available for download at **www.educationendowmentfoundation.org.uk**