Education Endowment Foundation

Quest
Evaluation report and Executive summary
June 2015

Queen's University Belfast

SCHOOL OF EDUCATION
The Centre for Effective Education

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 evaluation was conducted by the Centre for Effective Education (CEE), Queen's University Belfast. Support was provided by the Institute for Effective Education (IEE), University of York. As the IEE have established links with the programme developer, the CEE was appointed to independently oversee the evaluation, conduct the random allocation, analyse the data and produce an independent report. The main fieldwork was organised and collected by the Institute for Effective Education.

The lead evaluator was Dr Andy Biggart.

Institute for Effective Education

Contact details:

Dr Andy Biggart
Centre for Effective Education
School of Education
Queen's University Belfast
69–71 University Street
BT7 1HL

p: 028 9097 5946
e: a.biggart@qub.ac.uk

Contents

Executive summary

The project

Quest is a whole-year group approach to teaching English in Key Stage 3. Pupils are grouped according to their current level of reading comprehension, typically with smaller classes for the lowest attaining groups. Key components of the programme include: an emphasis on collaborative (or 'co-operative') learning; a requirement that participating teachers follow a consistent 'cycle of instruction'; and the use of formative assessment in every lesson. Pupil progress is reviewed every eight weeks, with results used as the basis for re-grouping the class.

The programme was designed by the charity Success for All and had been adapted from an existing programme available in the United States called Reading Edge.

This project sought to assess the impact of Quest on Year 7 pupils' reading comprehension and was supplemented by a process evaluation that assessed programme implementation. Six schools followed the programme, and a further 13 participated in the evaluation. All participating teachers received two days training and extensive resources to be used as part of the approach, including lesson plans, graded texts and 'digitexts' (interactive texts). The programme was designed to be delivered in daily 60-minute lessons by all Year 7 English teachers, and run over the full school year.

The study was funded by the Education Endowment Foundation as one of 24 projects in a themed round on literacy catch-up at the primary-secondary transition.

Key conclusions
1. The evaluation was unable to provide a secure estimate of the impact of Quest on reading comprehension outcomes among Year 7 pupils, primarily due to a high level of drop-out from the trial.
2. Few, if any, schools implemented the programme as designed, suggesting that substantial adaptation may be required if the approach is to gain wider adoption in English schools.
3. Some of the main barriers to successful implementation included: the difficulty in covering the expected material in a single school lesson; a perception of an insufficient focus on writing activities; and its adoption as a whole-year group intervention.
4. Many schools were positive about the range of resources provided by the programme and the co-operative learning aspects of the programme.
5. Though it is not possible to draw a conclusive statement about the impact of the programme, on average pupils who received the programme made less progress than those who did not.

Security rating

> *Security rating awarded as part of the EEF peer review process*

Findings from this trial have **low security**. The trial was set up as a randomised controlled trial that aimed to compare the progress of pupils who received the programme to similar pupils who did not. The trial was classified as an 'efficacy trial', meaning that it sought to test whether the intervention could work under ideal or developer-led conditions in ten or more schools. However, a large number of participating schools dropped out of the project which substantially reduced the security of the resulting impact estimates. In addition, there was some imbalance between the intervention and comparison groups in terms of the proportion of pupils eligible for free school meals and their gender.

There were high levels of attrition and missing data, particularly among the intervention group (from which 4 out of 13 schools withdrew, compared to two comparison schools). The high drop-out rate makes it hard to attribute any difference between intervention and comparison schools to the programme, rather than to chance. It is also possible that the level of drop-out may have reduced the accuracy of the estimate in other ways. For example, if those who dropped out of the programme were on average less engaged, the estimate could overstate the impact of the approach.

Results

- On average, pupils who received the programme made less progress than those who did not. However, due to the high dropout rate and the small sample size, it was not possible to draw a conclusive statement about the impact of the programme.
- The level of implementation of the programme in Quest schools was low, which may have diluted any potential impact on reading comprehension.
- Conversely, if those who dropped out of the programme were on average less engaged, or implemented the programme with less fidelity, the estimate could overstate the impact of the approach.
- The programme did not appear to have differential effects according to gender, free school meal eligibility, or prior reading attainment.
- The process evaluation highlighted a number of issues that both undermined teacher confidence in the programme and impacted upon the extent of implementation.
- The perceived weaknesses of Quest according to teachers were: that it was overly prescriptive; that the lesson plans contained too much material to be covered in a standard 50 or 60-minute school period; and that it contained insufficient writing opportunities.
- Positive feedback was provided by the teachers in relation to the range of books and digital texts provided for pupils, although these could be improved for 'lower ability' sets.
- On the whole, teachers were also positive about Quest's co-operative learning structure (particularly the teamwork and opportunities for discussion).

Cost

The cost of the approach as delivered in the trial is estimated at £161 per pupil. This estimate is based on a per-school cost of £24,171 and an average of 150 pupils per year group. The cost estimate includes costs for training, ongoing support for teachers, all programme materials and resources. It does not include the salary costs of participating teachers, or of any supply cover required to permit teachers to attend training.

Group	Effect size	Estimated months' progress	Security rating	Cost
Quest vs. comparison	-0.04	-1 months[1]	🔒🔒🔒🔒🔒	££

---

[1] *Since this report was published, the conversion from effect size into months of additional progress has been slightly revised. If this result was reported using the new conversion, it would be reported as 0 months of additional progress rather than -1. See **here** for more details.*

Introduction

1.1 Intervention

This is a report on an efficacy trial of the Quest programme. Quest is a whole-year group intervention that is designed to help struggling readers in early secondary school improve their literacy. Pupils are grouped according to their reading skills, with smaller classes for those with lower attainment. Teaching includes an emphasis on co-operative learning, with the most struggling pupils also receiving computer tutoring and individualised teaching.

The programme is designed to be delivered by all Year 7 English teachers in a school, and runs over a full school year (September to July). Throughout the school year pupils are reassessed at regular intervals and thus have the opportunity to move between ability groups as required. Another teacher within the school, ideally an English teacher who is not directly involved in the delivery of the programme, is appointed to act as the Quest facilitator. Their role is to coordinate and monitor the delivery of Quest across the whole year group. This typically includes regularly observing lessons, overseeing the timetabling of English lessons, pupil grouping by attainment, and regular assessments and regrouping.

1.2 Background evidence

Quest has evolved out of a project called The Reading Edge that has been previously used in the United States. The Reading Edge is a literacy programme developed for middle-grade students (aged 11–14 years), with the goal of preparing them to be strategic, independent and motivated readers (Slavin et al., 2009). It emphasises the use of co-operative learning strategies, for which teachers receive training and coaching. Co-operative learning is a learning approach in which students help each other to learn by working together in small groups on a structured learning task (Cooper et al., 1990; Slavin et al., 2009). The theory behind this approach is that students will learn more when they work together and effectively 'teach' each other, as opposed to traditional learning through whole class instruction. Co-operative learning has been the subject of several reviews and meta-analyses (Johnson et al., 1981; Slavin et al., 1983; Qin et al., 1995; Lou et al., 1996; Thanh et al., 2008; Hattie et al., 2009; Kyndt et al., 2013). These have generally shown that co-operative learning has a positive effect on student achievement.

The Reading Edge has been evaluated in the US using a randomised controlled trial design involving 788 sixth grade (11–12-year-old) students in high poverty schools. The evaluation showed a moderate improvement in overall reading test scores among children who received the programme, and in vocabulary and comprehension (Slavin et al., 2009). The Reading Edge has recently been adapted to the UK setting by Success for All (UK), however, this represents the first evaluation in an English context. The purpose of the current evaluation was to determine the impact of Quest in improving literacy outcomes among Year 7 pupils. This evaluation is set up as an efficacy trial. Efficacy trials aim to test whether an intervention can work under ideal conditions (for example, when being delivered by the intervention's original developer) in more than 10 schools.

The evaluation took place between September 2013 and June 2014 and comprised both an impact assessment and a process evaluation that involved both qualitative and survey methods of data collection.

1.3 Evaluation objectives

The purpose of the current evaluation was to determine the impact of Quest in improving reading comprehension among Year 7 pupils through a randomised controlled trial that was designed to compare the programme to control schools who continued with their normal literacy practice.

In addition to the outcome evaluation, a process evaluation was intended to assess issues in relation to the tutors' and pupils' experience of the programme, and to ascertain whether it could be effectively implemented in secondary schools in England.

It was funded as part of a series of evaluations through funding provided to the EEF by the Department for Education focused upon literacy catch-up programmes for pupils at the transition between primary and secondary school.

1.4 Project team

The evaluation team comprised researchers from the Centre for Effective Education (CEE) at Queen's University Belfast. CEE researchers were responsible for the design, randomisation, analysis and reporting of the results of the impact evaluation. Researchers from Institute for Effective Education (IEE) at the University of York worked alongside Success for All (SfA) to recruit the schools, as well as managing data collection at post-test. The majority of the process evaluation data was collected by the IEE research team, supplemented by a light-touch process evaluation by the independent evaluators at the CEE.

Independent evaluation team from Centre for Effective Education (CEE)

- Dr Andy Biggart, Deputy Director
- Dr Seaneen Sloan, Research Fellow
- Dr Liam O'Hare, Senior Research Fellow
- Dr Sarah Miller, Deputy Director at CEE

Institute for Effective Education (IEE)

- Professor Bette Chambers, Institute Director
- Dr Pam Hanley, Research Fellow
- Dr Sarah Blower, Research Associate

Success for All Foundation, UK

1.5 Ethical review

Ethical approval was granted by the Department of Education Research Ethics Committee, University of York in April 2013, and by the Research Ethics Committee, School of Education, Queen's University Belfast, in May 2014, prior to any data collection being undertaken. Informed consent was obtained at the school level from headteachers, and by individual teachers for interviews and classroom observations. Written information and opt-out consent forms were sent home to parents of eligible pupils.

1.6 Trial registration

The trial was registered with the ISRCTN registry (ref: ISRCTN78027416).

Methodology

2.1 Trial Design

The evaluation was designed as an efficacy trial to test whether Quest can work under ideal conditions. There were two trial arms: control and intervention. As the programme was designed to be delivered across a whole year group, randomisation was at the school level, and the Quest programme was delivered with Year 7 pupils (across the whole year group) in schools that were randomly allocated to the intervention group. Year 7 pupils in the control group were taught the English curriculum using standard practice. This design was chosen to provide a robust evaluation of the Quest programme.

Randomisation was conducted independently by the CEE following recruitment. In order reduce the amount of variation between schools in the intervention and control group, minimisation was used to conduct the randomisation process. This accounted for a number of school-level factors (see Section 2.5). The analysis was conducted to take account of clustering using robust standard errors.

It was originally intended that 30 schools would be recruited to the trial (see Sample Size section for more details), however, some difficulties were experienced with school recruitment (as schools were unwilling to make changes to the planned curriculum for Year 7). Ultimately, 25 schools were recruited to the study.

2.2 Eligibility

Schools were recruited throughout England, and although the majority was based in the North (North East and Yorkshire and Humber) other schools were recruited from the South East, South West and East regions. Schools were targeted on the basis of having a higher than average of proportion of pupils receiving free school meals ('FSM' pupils) and according to the following criteria:

- willing to be randomly assigned to either arm of the trial at the school level;
- willing to engage with the intervention and implement it with all Year 7 English classes; and
- willing to administer literacy tests at the end of the school year.

School-level consent from the headteacher was sought prior to randomisation. Opt-out parental consent for participating in the trial was also sought through schools (see Appendix A for parent information and consent forms).

2.3 Intervention

Quest is a year-long Key Stage 3 literacy programme that is aligned to the English Curriculum, from level 1 up to level 6 for reading and writing. In the intervention schools it was delivered by all English teachers who were responsible for teaching a Year 7 class. The programme was designed to be delivered daily during a 60-minute English lesson.

English teachers undertake two full days of training in preparation for delivering Quest. Training is provided by the SfA Foundation.

The key elements of Quest are:

Co-operative learning: pupils work in small, mixed ability learning teams under a set of instructional methods. Pupils read in pairs and practice reading together, with each of the partners taking turns to read a portion of the text. Team discussion is encouraged, and pupils are rewarded for their achievements as a team. Pupils in teams therefore take responsibility for helping their teammates to learn.

Instructional design: Each lesson incorporates the cycle of effective instruction, including active instruction, modelling and guided practice.

Grouping: Pupils are streamed according to their reading comprehension level. Pupils experience a variety of text types and genres selected on the basis of their comprehension level; these include fiction, non-fiction, media, web based, digitexts (online or digital interactive texts), poetry and graphic novels. Each Quest level incorporates reading and writing lessons with a clear focus on the pedagogical process and the continual assessment of student progress using explicit objectives and techniques to measure those outcomes. The most struggling pupils also receive computer assisted tutoring and individualised teaching.

Assessment: Formative assessment opportunities are integrated into all lessons, and pupils are regularly assessed on their fluency, vocabulary and writing. Progress is reviewed every eight weeks through summative assessments, and re-grouping may occur on the basis of this.

Control condition

The control condition was 'business as usual'. Teachers in the control group continued to teach the English curriculum using their normal methods. Schools in the control group received a payment of £1,500 for agreeing to take part in the evaluation.

2.4 Outcomes

As the project was funded as part of the Transitions Round, the primary outcome was pre-specified as reading comprehension (age-standardised score), as measured by the GL Assessment New Group Reading Test ('NGRT', digital edition).[2] This test is based on assessments of sentence and passage completion, and was administered at post-test. Schools administered the digital testing online under exam conditions and these were automatically scored by GL Assessment. It was not possible for schools to be blind to treatment allocation, however GL Assessment were not aware of allocation.

The KS2 reading score (standardised national school literacy assessment, sat by all pupils in Year 6 in May of 2014 and accessed through the National Pupil Database) acted as a pre-test measure of prior pupil attainment in the analysis models.

2.5 Sample size

At the design stage, the following parameters were used to calculate the required sample size:[3]

- Significance level (α) = 0.05
- Power (P) = 80%
- Site size (n) = 80
- Effect size (δ) = 0.2 and 0.3
- Intra-Cluster Correlation (ρ) = 0.10
- Proportion of variation at level 2 ($R_{2|2}$) = 0.60

Power calculations suggested that with 30 schools (and on average, 80 pupils per school), the initial study design has a power of 80% to detect an effect size of 0.23.

---

[2] NGRT information available at www.gl-assessment.co.uk/products/new-group-reading-test.
[3] Calculated using Optimal Design software (http://hlmsoft.net.od/)

2.5 Randomisation

After schools had been recruited by Success for All (UK) and the IEE, randomisation was conducted at the school level by the independent evaluation team at CEE. The CEE was initially asked to randomise a first batch of 20 schools for the trial, with the possibly of a further 10 to follow.

With randomising only 20 schools there was a risk of getting unbalanced samples by chance on key covariates if simple randomisation was used. This might have been exacerbated by a further randomisation of 10 or less schools. Minimisation was therefore adopted as the preferred method of allocation in the current study. Minimisation is a well-recognised approach that uses algorithms to ensure balance between certain covariates of control and intervention schools at baseline, and is especially useful when randomising a small number of cases (Torgerson and Torgerson, 2007). It was therefore used to ensure the groups would be as evenly matched as possible and was based on the four school-level characteristics (described below). Median values were examined for each of these characteristics to determine a mid-cut point.

The four school-level characteristics were:

School size: schools with 150 pupils or less were coded as '1', and schools with 151+ pupils were coded as '2'.

Proportion of FSM pupils: schools with less than 24% of pupils eligible for Free School Meals were coded as '1', and those with more than 24% eligible were coded as '2'.

Proportion of pupils with English as an Additional Language (EAL): schools with less than 5% EAL were coded as '1', and those with over 5% were coded as '2'.

Attainment: data on average GCSE passes at grades A*–C were available for the majority of schools, however two schools were middle schools and therefore KS2 data was the only available indicator of attainment. To create a common measure of attainment for all schools, the national average results were subtracted from the available score (GCSE or KS2) to indicate whether a school was above or below the national average. Median values were assessed and a value of -10% or below was coded as '1' and above this figure was coded as '2'.

These four dichotomous variables were set up in the Minim software package;[4] all variables were given a weight of one with the exception of attainment which was double weighted as an important predictor of the outcomes of interest.

Schools were then given a random number between 0 and 1, and schools were then ordered from low to high on this variable. The school with the lowest random number was selected first and entered into Minim. This randomly allocated it to the intervention group. Each school in random number order was then entered until all 20 schools had been allocated.

When additional schools were recruited they were allocated individually whereby this previous allocation process was taken into account in balancing the samples between the control and intervention schools according to the key characteristics outlined above.

---

4 http://www-users.york.ac.uk/~mb55/guide/minim.htm

2.7 Analysis

Initially, as the evaluation was designed as a cluster-randomised controlled trial with allocation at the school level, it was proposed that multi-level modelling would be used for analysis. However, due to the smaller than expected number of schools recruited to the study, it was agreed between the IEE and the CEE (in consultation with the EEF), prior to analysis, that the main outcomes would be analysed using regression with robust standard errors. This adjusts for the clustered nature of the allocation.

Binary dummy variables were used to identify the intervention (coded '1') and control (coded '0') groups. Binary dummy variables were also used for FSM eligibility (coded '1' for yes and '0' for no) and gender (coded '1' for male and '0' for female). The four binary school-level variables used in the randomisation process—school size, proportion of FSM pupils, proportion of EAL pupils, and attainment—were also included as covariates.

Analyses were conducted in Stata version 12 (Stata Corporation, College Station, Texas, USA) and on an intention-to-treat basis meaning that participants were analysed in the group to which they were randomised, irrespective of whether or not they actually implemented the programme. Analysis was conducted in stages: the main effects analysis was carried out first, followed by pre-specified sub-group analyses examining interaction effects. The main effects models included a dummy variable for the intervention status, and controlled for prior reading attainment (K2) as a pre-test measure and included variables for gender, age and FSM status as additional covariates. The exploratory analyses examined whether the effect of the intervention varied between different subgroups of participants—boys and girls, FSM eligibility, and prior (KS2) reading attainment. Interaction terms based upon each of these variables, as well as whether or not in the Quest group, were then added separately to the models.

Effect size (Hedges' g) was calculated as the standardised mean difference between the control and intervention groups, using the pooled standard deviation. The pooled standard deviation was calculated using the formula:

$$s = \frac{\sqrt{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}}{n_1 + n_2 - 2}$$

Hedges' g was then calculated as:

$$g = \frac{\text{coefficient}}{\text{pooled standard deviation}}$$

Pre-specified exploratory analysis was also undertaken to determine whether the quality of programme delivery (in terms of fidelity to the Quest programme) was associated with the outcome. Fidelity rating was undertaken by the IEE during lesson observations (see Section 2.8, below).

2.8 Process evaluation methodology

Table 3.1: Description of process evaluation data collection

Date	Activity
September 2013–June 2014	Lesson observations (IEE)
June–July 2014	Quest facilitator interviews (IEE)
June–July 2014	Teacher and student online survey (IEE)
June–July 2014	Telephone interviews with teachers (CEE)
August–November 2014	Process evaluation analysis and write up (IEE and CEE)

All schools in the intervention group were invited to take part in each aspect of the process evaluation, and all schools in both the control and intervention groups were invited to complete the teacher and student online surveys.

Lesson observations were undertaken and analysed by the IEE in eight schools who were delivering Quest (this included one school who later withdrew). A total of 13 lessons were observed in the eight schools. Each lesson was given an overall implementation rating ranging from 0 to 3 (see Appendix C for full fidelity ratings). This rating formed a quantitative measure of fidelity, with higher scores reflected greater fidelity to the programme (that is, adherence to Quest routines, use of materials and consistent use of co-operative learning).

All schools were sent the survey link (via the key contact) and asked to circulate it to all teachers taking Year 7 English. The survey asked teachers how well their current English scheme supported various skills and components (such as writing, reading fluency, comprehension, grammar, language acquisition, independent work, collaborative work, monitoring and assessment). Teachers were also asked to rate their attitude towards teaching Year 7 English, in terms of their own confidence, classroom management skills, enjoyment of teaching and perceived pupil progress. Similarly, the student survey was completed online and all Year 7 pupils were invited to participate. The student survey covered their attitude towards and enjoyment of English lessons, self-perception of ability and progress in English, and how often different teaching methods are used (such as partner or team work, discussions, or working in silence). Students in the intervention group were also asked how much they enjoyed Quest lessons compared with other English lessons.

The Quest facilitator interviews and the online teacher and student surveys were undertaken and analysed by the IEE, with a written summary provided to the CEE. The CEE contacted the Quest facilitator at all intervention schools and asked for an invitation to be sent to all Year 7 English teachers to take part in a telephone interview. Interviews lasted, on average, 30 minutes and covered topics such as: satisfaction with the amount of training and support for delivering the programme; perception of materials and resources provided; any challenges or successes associated with delivering the programme; any adaptations that had been made to the programme; benefits for students; and impact on their own teaching role.

Costs

Information on the cost of the programme—including the cost of training and ongoing support as well as programme materials—was obtained from the programme developer, Success for All (UK). Cost per pupil was calculated by dividing the cost per school by an average year group size of 150 pupils.

Impact evaluation

3.1 Timeline

Table 3.2: Timeline of activities related to the trial

Date	Activity
February–July 2013	Recruitment (SFA)
June 2013 (and July 2013)	Randomisation (CEE)
September 2013	Teacher training (SfA)
September 2013–July 2014	Programme delivery (teachers supported by SfA)
June–July 2014	Post-test data collection (IEE)
October 2014	Pre-test data (KS2) made available by NPD
December 2014	Analysis and write-up of impact evaluation (CEE)

3.2 Participants

Recruitment was undertaken by both Success for All (UK) and the IEE between February and July 2013. In total, 267 secondary schools were contacted in writing and invited to take part in the trial. Schools were recruited throughout England, and although the majority were based in the North (North East and Yorkshire and Humber) other schools were recruited from the South East, South West and East regions.

There were some difficulties with recruitment as schools were reluctant to change their planned delivery of the Year 7 English curriculum. While schools were initially interested in the programme, when they found out it would involve a complete change of approach to their planned Year 7 English curriculum many declined further participation. Others had hoped they could use the programme as a targeted intervention for those who were most struggling with English and declined when they realised that they would need to implement the programme across the whole of Year 7. In total, 61 schools assessed for eligibility subsequently decided not to take part in the evaluation. As a result, fewer schools than originally planned were recruited to the study. Initially, 25 schools agreed to take part, and of these, 13 were randomly allocated to receive Quest and 12 allocated to the control group. However, 8 of the 25 schools withdrew after being made aware of their allocation (see Figure 3.1). Six of these schools were from the intervention group, and two were from the control group. Two of the schools in the intervention group subsequently withdrew because the programme conflicted with planned Year 7 provision. Another four intervention schools withdrew at the beginning of the autumn term: one school was in the process of closing and pupils were being relocated to other schools; two schools were dissatisfied with the programme; and one school decided it was not in a position to begin implementing Quest fully until the spring term.

Significant efforts were made in the case of school withdrawals to persuade schools to remain in the evaluation and to provide post-test data for the intention-to-treat analysis. However, only three of the six withdrawn intervention schools provided post-test scores for inclusion in the analysis. One further school in the intervention group (not withdrawn) did not provide any post-test data and was therefore not included in analysis. Despite control schools signing a memorandum of understanding and the offer of a financial incentive for completing post-tests, two control schools failed to respond to repeated requests to arrange testing. In total, four intervention schools (n = 589 pupils) and two control schools (n = 220 pupils) were lost to follow-up. Post-test data was provided by 19 schools (9 intervention schools and 10 control schools).

3.3 Incomplete outcome data

Some outcome data was missing from the 19 schools that provided post-test data, particularly within intervention schools where 34% of pupils were not post-tested compared to 15% of pupils in the control group (Table 3.3). As schools were not supervised during the online testing, one plausible reason for the discrepancy is that some schools only tested those pupils from classes who had been more directly involved with Quest. Some schools also reported technical problems with digital tests as a reason behind incomplete testing such as difficulty in accessing tests (due to the expiry of log-in details which then had to be reissued) and difficulties in contacting test supplier technical support. Some students received an error notice on completion of the test which resulted in no test score being returned. However, these latter reasons would not easily explain the extent of the discrepancy observed between the control and intervention schools.

Table 3.3: Proportion of missing post-test data within schools[1]

	Intervention group (9 schools)	Control group (10 schools)	Total
Number of pupils with post-test data	959 (66%)	1,158 (85%)	2,117
Number of pupils with no post-test data	502 (34%)	213 (15%)	715
Total	1,461	1,371	2,832

[1]Pearson chi-square test (df) = 132.79 (1), $p < 0.001$

There was no relationship between being post-tested or not in the case of both gender and FSM eligibility. Pupils who were not posted-tested, however, scored lower in KS2 reading (mean points = 26.1, sd = 6.27) compared to those post-tested (mean points = 27.4, sd = 5.19, $p < 0.001$).

Figure 3.1: Trial participants
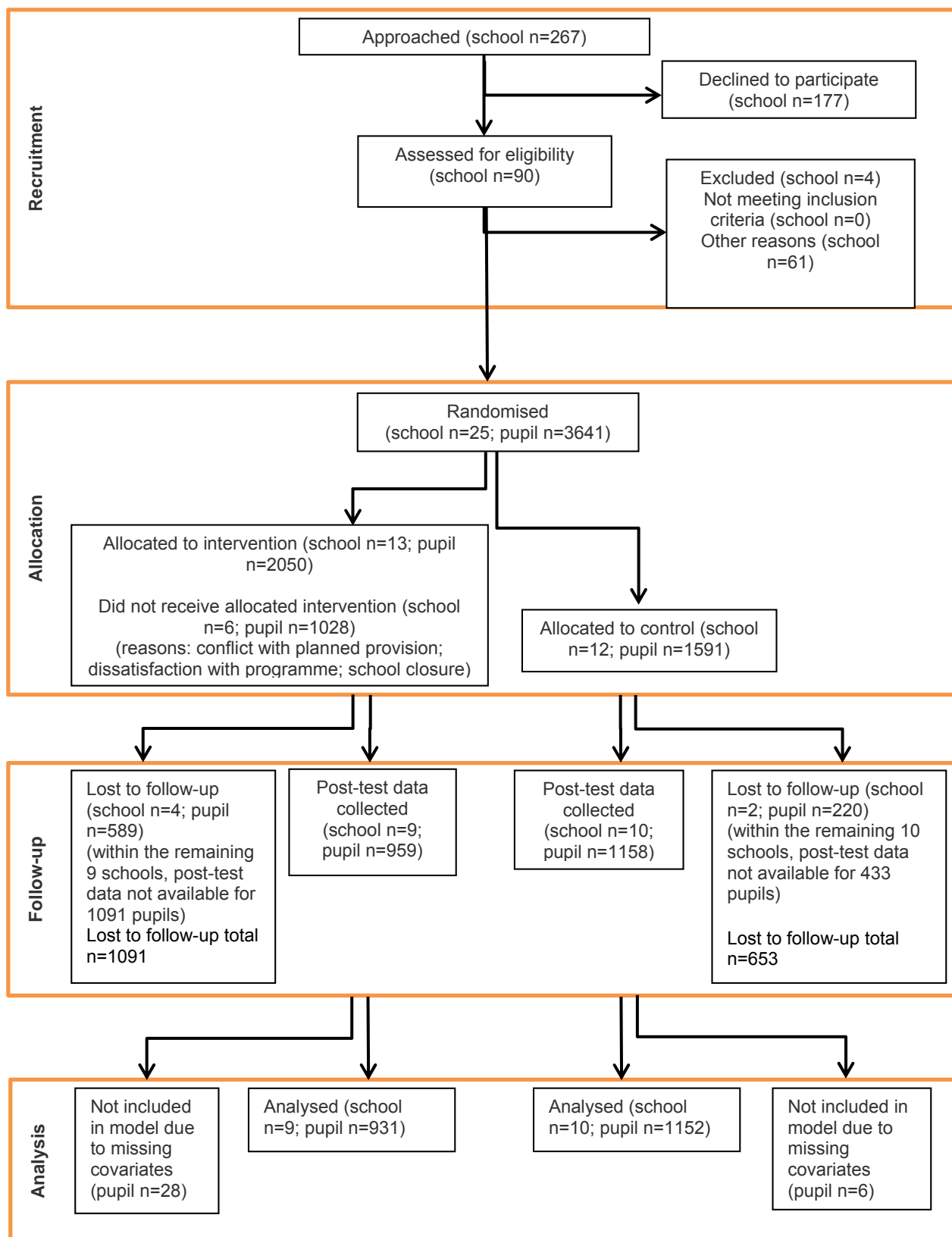
3.3 School characteristics

Compared to national figures, and in line with the recruitment strategy, schools recruited to the trial had a higher percentage of FSM pupils, a higher percentage of EAL pupils, and a slightly higher percentage of pupils with a statement of Special Educational Needs (SEN) (Table 3.4). Schools in the sample had a lower percentage of pupils achieving 5+ GCSEs (including English and mathematics) at grades A*–C compared to national data. Overall, there was little difference in the original sample (of 25 schools) recruited to the study and the 19 schools included in the analysis, although the proportion of pupils with EAL was slightly higher in the original sample.

The majority of schools recruited were Foundation schools (n = 9), followed by Community Schools (n = 6). Five schools were Academy (sponsor-led) schools, and four were Academy converter schools. One school was voluntary-aided. Of the five schools that withdrew, two were Foundation schools, two were Community schools, and one was voluntary-aided (Table 3.5).

Table 3.4: Characteristics of sampled schools compared with national data

	England—national (secondary state-funded)	Original sample (n=25)	Sample for analysis (n=19)
Percentage of pupils with SEN	1.9%	2.5%	3.0%
Percentage of pupils with English as an Additional Language (EAL)	13.6%	18.2%	14.2%
Percentage of pupils eligible for Free School Meals (FSM)	16.3%	25.6%	26.1%
Percentage of pupils achieving 5+ A*–C or equivalent (including both English and Mathematics)	60.6%	47.2%	47.8%

The majority of schools were based in an urban setting; 4 out of 25 schools (or 3 out of the 19 in the sample for analysis) were located in a rural setting (Table 3.5).

Nine of the recruited schools were rated as 'Good' by Ofsted; two of these schools withdrew (from the control group). Seven of the recruited schools were rated as 'Requires improvement', and four were rated as 'Inadequate'. Three of the four schools rated as inadequate did not provide post-test data (Table 3.5).

Table 3.5: School characteristics

	Original sample (n=25)		Sample for analysis (n=19)	
	Intervention	Control	Intervention	Control
School type				
Foundation School	4	5	2	4
Community School	3	3	1	3
Academy—Sponsor-Led	4	1	4	1
Academy Mainstream Converter	2	2	2	2
Voluntary Aided School	0	1	0	0
School location				
Urban	10	11	7	9
Rural	3	1	2	1
Ofsted rating*				
Good	3	6	3	4
Requires improvement	5	2	4	2
Inadequate	3	1	1	0

Ofsted rating not available for five schools.

3.4 Pupil characteristics

Pupil characteristics are displayed in Table 3.6. The randomisation process resulted in intervention and control groups that were comparable at baseline in terms of prior attainment (KS2 reading scores). The control group had a higher proportion of girls, and a higher proportion of FSM pupils, compared to the intervention group. A similar pattern was reflected in the sample for analysis.

Table 3.6: Pupil characteristics

	Original sample (n=3641)		Sample for analysis (n=2117)	
	Intervention (n=2050)	Control (n=1591)	Intervention (n=959)	Control (n=1158)
Gender Male Female *missing*	1053 (51%) 951 (46%) *46*	686 (43%) 885 (56%) *20*	524 (56%) 412 (44%) *23*	513 (44%) 642 (56%) *3*
Pearson chi-square test	Value (d.f)= 27.79 (1)	$p<0.001$	Value (d.f)= 27.67 (1)	$p<0.001$
FSM eligible No Yes *missing*	1249 (61%) 775 (38%) *46*	887 (56%) 694 (44%) *20*	602 (64%) 334 (36%) *23*	664 (57%) 491 (43%) *3*
Pearson chi-square test	Value (d.f)= 15.55 (1)	$p<0.001$	Value (d.f)= 10.09 (1)	$p=0.001$
KS2 Reading score (pre-test) Mean (sd) Median (IQR) No. of obs *Missing*	27.02 (5.64) 28.26 (6.5) 2008 *42*	27.03 (5.34) 28.26 (5.7) 1564 *27*	27.36 (5.44) 28.26 (6.1) 938 *21*	27.49 (4.98) 28.56 (5.8) 1152 *6*
t-test	mean difference (d.f.)= 0.045 (3570)	$p=0.96$	mean difference (d.f.)= 0.564 (2088)	$p=0.57$

Table 3.6: Pupil characteristics

	Sample for analysis (n=2117)	
	Intervention (n=959)	Control (n=1158)
Gender		
Male	524 (56%)	513 (44%)
Female	412 (44%)	642 (56%)
missing	*23*	*3*
FSM eligible		
No	602 (64%)	664 (57%)
Yes	334 (36%)	491 (43%)
missing	*23*	*3*
KS2 Reading score (pre-test)		
Mean (sd)	27.36 (5.44)	27.49 (4.98)
Median (IQR)	28.26 (6.1)	28.56 (5.8)
No. of obs	938	1152
Missing	*21*	*6*

3.5 Outcomes and analysis

The aim of the impact evaluation is to provide initial quantitative data on the efficacy of Quest in terms of improving reading outcomes for Year 7 pupils. The main analysis was undertaken using regression, controlling for the effects of pupil age, gender and FSM eligibility, as well as taking into account any clustering at the school level through robust standard errors. The outcome variable (reading comprehension, measured by the NGRT) and covariates were standardised prior to analysis. Effect sizes were calculated using Hedges' *g*. Pre-specified subgroup analysis were also undertaken to explore any differential effects according to pupil prior attainment (at KS2), gender, and FSM eligibility.

Table 3.7 displays the results of the main analysis, and the results of the subgroup analysis are shown in Table 3.8. In both of these tables, the coefficient refers to the mean change in the dependent variable (in other words the outcome—reading comprehension) for one unit change in the predictor variable (such as Group), while holding the other predictors in the model constant. Coefficients can be positive, zero, or negative. A positive coefficient indicates a higher score on the outcome variable, for example, if a positive coefficient is seen for the 'Group' variable, then it indicates that those in the intervention group score higher than those in the control group. Similarly, a negative score indicates a lower score on the outcome variable, and a coefficient of zero indicates no difference in the outcome variable. The standard error associated with each coefficient is also reported, and this gives an indication of how precise the estimated coefficient term is (with small values reflecting greater precision). Statistically significant coefficients are indicated by a p-value of less than 0.05, meaning that the finding is unlikely to have been caused by random error—by chance.

Table 3.9 displays the adjusted post-test means, and the effect size and 95% confidence intervals for the main effects model and also for the subgroup of pupils eligible for free school meals.

Table 3.7: Main effects model for outcome—reading comprehension (NGRT)

	Main Effects Model		
	Coefficient	**Robust standard error**	**95% Confidence Interval**
Group (Intervention)	-0.038	0.05	-0.14, 0.07
Pupil-level variables			
Pre-test (KS2 reading)	0.73***	0.04	0.65, 0.82
Gender (male)	-0.08***	0.02	-0.12, -0.05
FSM eligibility	-0.07***	0.01	-0.10, -0.04
Age at post-test	-0.07***	0.02	-0.10, -0.04
School-level variables			
School size (above average)	-0.01	0.04	-0.09, 0.07
Proportion FSM (above average)	0.06	0.03	-0.01, 0.12
Proportion EAL (above average)	-0.02	0.02	-0.07, 0.04
Attainment (above average)	0.09**	0.03	0.03, 0.15
Constant	-0.03	0.03	
No. of observations	2083		

*Significance: $p < 0.05$ **$p < 0.01$ ***$p < 0.001$*

Table 3.8: Interaction models for outcome—reading comprehension (NGRT)

	Interaction Models					
	Pre-test		**Gender**		**FSM**	
	coef.	**s.e.**	**coef.**	**s.e.**	**coef.**	**s.e.**
Group (Intervention)	-0.04	0.05	-0.04	0.05	-0.04	0.05
Pre-test (KS2 reading)	0.73***	0.03	0.73***	0.04	0.73***	0.04
Interaction (Pre-test × Intervention)	0.02	0.08	-	-	-	-
Gender (male)	-0.08***	0.02	-0.07*	0.03	-0.08***	0.02
Interaction (Gender × Intervention)	-	-	-0.03	0.03	-	-
FSM eligibility	-0.07***	0.01	-0.07***	0.01	-0.05**	0.02
Interaction (FSM × Intervention)	-	-	-	-	-0.05	0.03
Age at post-test	-0.07***	0.02	-0.07***	0.02	-0.07***	0.02
School-level variables						
School size (above average)	-0.01	0.04	-0.01	0.04	-0.01	0.04
Proportion FSM (above average)	0.06	0.03	0.05	0.03	0.06	0.03
Proportion EAL (above average)	-0.01	0.02	-0.01	0.02	-0.01	0.02
Attainment (above average)	0.09**	0.03	0.09**	0.03	0.08**	0.03
Constant	-0.03	0.03	-0.03	0.03	-0.03	0.03
No. of observations	2083		2083		2083	

*$p < 0.05$ **$p < 0.01$ ***$p < 0.001$

Table 3.9: Summary of main effects at post-test

	Adjusted post-test means				Effect size		
	Intervention group		**Control group**				
Outcome	**n (missing)**	**Mean (sd)**	**n (missing)**	**Mean (sd)**	**n in model (intervention; control)**	**Hedges g** (95% CI)	**p-value**
Reading comprehension	*959(1091)*	*-0.06 (1.06)*	*1158 (433)*	*-0.03 (0.95)*	*2083 (931; 1152)*	*-0.04 (-0.14, 0.07)*	*0.48*
Reading comprehension (FSM only)	*357 (25)*	*-0.20 (1.00)*	*494 (4)*	*-0.10 (0.94)*	*822 (332; 490)*	*-0.11 (-0.24, 0.02)*	*0.13*

A small negative effect size was found but was not statistically significant. In addition, none of the interaction terms were significant. Given the lack of significant results and, more importantly, the high level of attrition in the sample which is likely to have biased any estimates, there is no evidence to suggest the programme had an impact on reading comprehension or evidence to suggest that the programme worked differently for subgroups of pupils, at least in terms of prior attainment, gender and free school meal eligibility.

On-treatment analysis

As described in the Process Evaluation Methodology, a number of lessons were observed in each school and rated for fidelity on a scale of 0 to 3, with higher ratings reflecting greater fidelity to the programme. The results suggest that schools with a low fidelity rating had improved outcomes relative to those who had not implemented the programme successfully at all (Table 3.10). However, this finding was not consistent across the two higher fidelity groups (Medium and High). As these ratings were based on single observations from a small number of ability groups in each school they may also not accurately reflect a valid measure of fidelity across the entire school, therefore any inferences drawn from these findings would need to be treated with caution.

Table 3.10: Effect of implementation fidelity on outcome

	Model (primary outcome—reading comprehension)	
	coef.	s.e.
Fidelity:[a]		
Rating Low	0.38***	0.06
Rating Medium	0.17	0.07
Rating High	-0.05	0.07
Pre-test (KS2 reading)	0.70***	0.10
Age at post-test	-0.07**	0.04
Gender	-0.10***	0.03
FSM	-0.08*	0.03
Constant	-0.11	0.02
No. of observations	931	

*$p < 0.05$**$p < 0.01$ ***$p < 0.001$

[a]*Reference Category: Fidelity rating not implemented.*

3.6 Cost

Quest costs £24,171 per school to implement with one full year group. This figure includes:

- 2 ½ training days (at £850 per unit) = £2,125
- 7 ½ support days (at £650 per unit) = £4,875
- teacher handbooks and student book = £6,500
- real texts = £10,671.

Additional and in-kind costs to schools include the additional non-class time of the teaching staff required to work with SfA to ensure the delivery of the programme, including the 2.5 training days. These may have to be covered by substitute teaching staff. Additional time also needs to be set aside for the member of teaching staff in each school that acts as the overall co-ordinator/facilitator for Quest.

Assuming an average of 150 pupils per year group, the average cost of the programme is approximately £161 per pupil. This cost is, however, likely to be an underestimate of the actual full economic cost of the programme which would also need to include estimates of the cost of teacher time. These are important to include as they represent significant 'opportunity costs' in that they are resources that could be used on other programmes. However, an estimate of the total cost of this programme would require a full economic cost-effectiveness analysis and was beyond the scope of this present study.

There was no charge to the schools involved in the current evaluation.

Process evaluation

This section presents the key findings from the process evaluation, bringing together the findings from the various components of the process evaluation—the lesson observations, the interviews with staff, and the teacher and student online survey.

Respondents came from five intervention schools (11 teachers) and ten control schools (35 teachers). They were mostly female (10 of 11 Quest teachers, 25 of 35 control). Most teachers in the control group had been teaching for 5 years or fewer (20/35), whereas the intervention teachers had more experience (9/11 teaching for 6+ years). The most common English qualification was a degree, although 2/11 intervention teachers only had GCSE English or equivalent.

The student survey was completed by 398 students from 5 intervention schools, and 779 students from 11 control schools. Just over half were girls (53%).

Telephone interviews with teachers were undertaken by the CEE as part of the fully independent aspects of the evaluation. All teachers in the seven schools who were still delivering the programme at the time of data collection (June–July 2014) were invited to participate through initial contact with the Quest facilitator in each school. Six teachers (in five schools) gave their consent to be contacted and were interviewed.

4.1 Implementation

Training and support for delivery

Teachers were asked about training and support as part of the online survey (to which 11 teachers in the intervention group responded). Most teachers (8 out of 11) considered the training they had received from SfA to be 'very' or 'quite' good, with two rating it 'not very good'. One teacher would have liked to have seen it modelled with a real Year 7 class (especially as regards the length), and another felt that ongoing training would have been valuable. Opinions on support from SfA were divided, with five rating it 'very' or 'quite' good, and five rating it 'not very' or 'not at all' good. The main issue was resources arriving late.

One facilitator that was interviewed felt that the initial training was patronising and not appropriate for the level of teaching experience of the audience. Facilitators reported that they had found it useful to have discussions about how they could adapt the programme to fit their particular circumstances, reflecting a feeling that adaptations to the programme were required.

One teacher felt that the training could have been improved by placing more emphasis on writing tasks, and another teacher commented that a demonstration of a full lesson would have been helpful:

'We didn't ever really get guidance as to how to get the whole of the Quest lesson into a lesson. Well, what we all kept asking 'Can you show us how to do a whole lesson,' even if it was just on a video, because we couldn't see how on earth to fit everything into a lesson.'

Acceptance of the programme

Programme resources

There was mixed feedback about the texts provided in the programme. The wide range of texts was appreciated and some of the individual texts were highly praised by both teachers and facilitators:

'I think the best bit of Quest is the absolute amazing variety of texts that the kids have access to.'

'Usually at Year 7 they'd see a variety of texts but not the massive variety that they've had access to on Quest, and they [the students] really enjoyed them, they got a lot out of it. So yeah, that part of it has been fantastic.'

However, there were some criticisms of some of the resources. Some teachers criticised some of the texts as inappropriate for the ability of the students, or having unappealing content. For example, a bias was noted towards the topic of football in the lowest ability books that proved a problem for a teacher whose group consisted of girls.

'I found the texts for my group [lower ability] to be a bit uninspiring. We have this issue all the time, because obviously they are low ability but they're not young children, so trying to find material that's fit for them is quite a difficult job, because you don't want to patronise them.'

One teacher was initially sceptical about using 'digitexts' and didn't expect the students to read from a laptop, but found that they really engaged the students:

'I wasn't sure of at all, because normally when you put young students onto a laptop to read, they don't really read. But the way they're designed with Quest, they're amazing and the kids were totally engaged. You could really take a step back and just let them do their learning really, because of the way the digitexts for Quest are designed. I'm really impressed with those.'

'It's [the digitexts] sort of designed so they can't just whizz through it, they have to follow the system and get all the information that they should be getting.'

One teacher praised the programme materials, although another felt that the materials needed further work in order to better link activities and objectives together:

'I think there were lots of aspects that were really good actually. The handouts were quite well put together and a lot of the materials were well done.'

'I found it quite fractured at times, because you'd have smaller tasks and then it didn't lead up to the bigger task, the bigger task was just put in there separately, so there was no lead-up really. So it would be better if the smaller tasks directly linked to the main task within the lesson, and the lessons were shorter as well, because the lessons had quite a lot of material in it, and you didn't always link to learning objectives.'

This was also reflected in feedback from another teacher who felt that some of the grammar elements were not well integrated into the context of the lesson, at times seeming random or standalone. The teacher felt these had been 'stuck on' to a lesson, perhaps because of recent changes to the English curriculum to incorporate spelling and grammar into lessons. The teacher had to adapt some of the materials to work the grammar elements more fully into the context of the lesson.

There was positive comment about some of the resources, such as the essay planning. One facilitator wanted more dynamic and editable material. However, the most serious criticism was that sometimes they were 'sloppily proof-read'—a particular problem with a literacy programme. Late delivery of books and other resources to schools was not uncommon, and this had caused difficulties with lesson planning. Both these issues appeared to be related to the early stage of development of the programme, with resources being produced to tight deadlines.

Co-operative learning

Facilitators and teachers were, on the whole, positive about the co-operative learning structure provided by Quest. Two of the schools already used co-operative learning techniques: one facilitator felt Quest had built on these, but the other thought it had diminished what they did:

"'Everyone was quite keen on the co-operative learning, although we did feel like it was revisiting a lot of old ground, and that sort of seemed to reinforce a lot of things that we were already doing which was good.'"

'I think in actual fact, some teachers felt that the level of co-operative learning in Quest was less than what they had been doing before.'

Although there was a recognition that it had been difficult to grasp and implement, all the other schools were positive about the co-operative learning. Some schools had even adopted the techniques with other year groups. They especially welcomed the group working, the competitive element and the opportunities for talk and discussion.

'They [students] love the competitiveness of the points. They love the fact they can talk. They like the fact there's not always loads and loads of writing … they love the discussion, the interactiveness [sic] of it. I don't think I've been in one of the classes where they've not been enjoying doing it.'

However, one teacher felt that this particular element was burdensome:

'I found it impossible to teach the lesson, keep an eye on everybody and also go around giving points. So from my point of view it was too much to do.'

Teachers were asked about their satisfaction with Quest as part of the online survey. Asked how they would prefer to teach Year 7 English the following year, three wanted to go back to how they used to teach it, two wanted to carry on with Quest, and another four intended to adopt a modified version of Quest (for instance, using it for only half the English lessons or using some of its methods for managing learning but not using the Quest materials). The main positives were that they liked the resources (3) and that students were more engaged (2), whereas the most common criticism was that there was too much content for one lesson (4). The main strengths of Quest were identified as the collaborative teamwork (4) and the texts (5).

Comments made by teachers through the online survey seemed to be quite polarised—it is possible that this simply reflected the nature of those teachers who had been motivated to complete the survey. One very enthusiastic teacher said: '*This scheme has real potential to be everything a school needs to bring about an engaging curriculum that provides ample opportunity for outstanding progress*' and had successfully used the pedagogy and ethos for Year 11 exam revision. In contrast, another '*found it took too long getting through all the tasks in one lesson consequently learning became disjointed; learning objectives weren't always specific.*'

Barriers to delivery

Logistical issues around grouping and regrouping students

One of the challenges to implementing the programme was the logistics around timetabling when students were split into different sized ability groups for daily English lessons. As one teacher said:

'Because we're a secondary school, we don't have those children in our class all day, so we were sharing classes and it was very problematic at the beginning to try and keep the continuity going.… it's quite difficult in a secondary school because it's very rare that the same teacher is going to be able to teach that same class every day. It just doesn't fit in with the rest of the timetable.'

Students are routinely reassessed and can be regrouped if necessary. Some teachers found that this could be problematic:

'Because it's so strictly set, they've got to move up and down, which, as a teacher of Year 7, I didn't really like because you need continuity with the younger children. And as a teacher you're able to cope with different abilities within a class, but [the problem is that] as soon as they've made enough progress they either go up, or if they weren't making enough progress they'd have to go down, so you lose your handle on those children.'

'Yeah, lots of movement between groups, which is then problematic for continuity, especially when children have just come up to secondary school. They didn't like to keep moving groups. Some children would just stay in their one set for a year.'

Overly prescriptive

A major criticism of Quest, reported by both teachers and facilitators, was the perceived rigidity of approach. Facilitators reported that some teachers found it too prescriptive, and wanted to alter it to better suit their students. As the year had progressed, it seemed that more adaptations were made, but in some cases this had been a main cause of Quest being dropped or downgraded in terms of lessons per week.

'So the two teachers that I [mentioned] that were really faithful to it, they quite liked it, having it all mapped out, knowing exactly what to do and they just pretty much followed it to the letter, and they seemed to like it. And I think their classes liked it well enough… She [another teacher] is a very creative teacher and I think that she found it quite hard sort of plodding through all the activities, when she would always have lots of ideas and want to think, "Oh, we could do this and I could bring this in." And so I think that she felt that had been limited.'

Lesson timing

Another problem was that Quest had been designed in lesson chunks that were longer than the fifty or sixty minutes slots available at most of the schools. Teachers had different ways of dealing with this, but most were left feeling dissatisfied by having too much content condensed into one lesson, or disjointed lessons if they chose to spread the content across two slots. They found themselves unable to finish the six-week units and only the most confident was radical enough to cut out a sufficient amount to fit the timeframe:

'…also I think people feel maybe a little bit demoralised by the lesson plans because they're trying to plan a lesson and going in with in the knowledge that they will never be able to get through everything in the unit. And it's like a psychological thing of "I've failed before I've even started". That's not something that I feel because I think you've got to be a professional and should pick and choose the bits that you think are the most important and then as you get better at it and quicker then you can introduce more things, but that's just something that other people have approached me with as a thought.'

This is consistent with the online teacher survey, in which four teachers felt the amount of material to fit in a lesson was excessive and the main weakness of the programme.

Two teachers described the impact of the discrepancy between timing for lesson plans and actual lesson duration:

'We couldn't fit it into a lesson. And then they told us actually they were 90 minute lessons rather than an hour lesson, so we were always sort of catching up on ourselves because you couldn't fit it all in.'

'It would often say "read chapters one to four", and we'd only have time to read one. So I would just carry it over. And it put the whole thing completely out of sync.'

Lack of focus on writing

Although acknowledging the link between reading and writing, teachers were uncomfortable with the limited amount of written work in Quest. They wanted more opportunities for extended writing practice, pointing out that this was something expected in English development and future exams:

> *'…we have adapted it so that it works for us, for our staff and for our children, because if you don't have as much writing, there is no evidence, in books or anywhere, that the children have made progress, and I think if people come in and say "Where is this? Where is that?", and you have just got a series of one or two words, or short paragraph answers, that is not preparing them for what they need to be able to do…'*

This issue was also reflected in some responses to the teacher survey; three teachers felt that a lack of focus on writing was the main weakness of Quest.

4.2 Fidelity

Six out of the 13 schools which had been allocated to the intervention group later withdrew from delivering the programme. Two of these schools withdrew soon after allocation because the programme conflicted with planned Year 7 provision. Two further schools withdrew because logistical reasons meant they were not able to deliver the programme. The remaining two schools withdrew because of dissatisfaction with the programme.

Thirteen sessions in total were observed across eight schools that were delivering the programme at the time of fieldwork. Each lesson was given an overall implementation fidelity rating, ranging from 0 to 3. The points on the scale were defined as follows:

Table 5.1: Fidelity ratings

Fidelity rating	Fidelity description
0	Quest in name only, no fidelity to the programme materials or co-operative learning.
1	Quest materials are in place, but the programme is not consistently/universally followed with fidelity. Co-operative learning is patchy.
2	Quest materials and routines are followed with fidelity. Co-operative learning is not consistent.
3	Quest materials and routines are followed with fidelity, and co-operative learning is embedded in the lessons.

Three of the lessons (delivered at three different schools) received the highest rating (Table B, Appendix B). Six lessons had an overall rating of '2', and four lessons (at two schools) had a rating of '1'. No lessons were rated as '0' in terms of implementation fidelity.

Ratings were also given to different aspects of the lesson, representing how often core features of the Quest approach were used. The highest scores received related to two criteria that also characterise well-delivered non-Quest lessons: 'the teacher circulates and supports students' and 'the teacher has students read books at students' level'. The frequency of use of other more Quest-specific features was low, such as: the award of team points; student 'think-pair-share'; and teachers modelling skills by using techniques such as 'think alouds'. When such features were introduced, students clearly recognised them and responded appropriately, leading to a relatively high rating for 'pupils are familiar with routines'.

Interviews with Quest facilitators in each school (whose role is to coordinate and monitor the delivery of the programme and manage the timetabling, differentiation, and assessment process), suggested

that there was considerable variation between and within schools in terms of delivery. For example, only one school reported doing Quest daily across all Year 7 classes. In another school, only two classes used Quest daily, with another class using it once a week, and the remaining three classes not using the programme at all.

One facilitator commented that their school reduced the amount of Quest that was delivered over time due to concerns about how much their students were benefitting from it:

'I think later there did become concerns that we weren't sure how well it was benefiting our students, and then once people started having those discussions, I think people felt sort of like a change in the culture and we felt a bit more negative towards it and a bit more worried about it, I suppose. Which was why it was then negotiated that we could reduce it right down.'

Issues around fidelity are reflected in the comments from teachers about how they felt they had to adapt the programme to best suit their needs:

> 'So I used the bits I effectively liked and the bits that worked and really didn't do much of, you know, the judge and jury and lots of other things that were on the plan. And I will admit, I wasn't the greatest proponent of Quest. I only taught… I didn't teach all of their lessons, I taught three a week, I think, and while I liked the concept, the idea of teamwork and a couple of the little tools they give you, I didn't follow it to the letter.'

> 'I used some of the resources… when we were doing narrative poetry, I'd got some ballad work of my own, so I used a little bit of that. But no, actually I adapted things a little bit. Sometimes with images, the images they provided, I thought, "well, you know what, I can find a better one than that". But actually resource-wise and text-wise it was all there, really, just a matter of going through it and picking which bits you could fit in.'

> 'We had to drop some of it [Quest content]… and that's when we had the refresher. When she [SfA trainer] came back she was horrified because I think a lot of us weren't even getting to the Quest questions, and she said they're the most important bits. So then as teachers, yeah, you had to sort of use your professional judgement as to which bits were important and which bits you could miss out.'

4.3 Outcomes of the programme

There was a mix of opinions on whether the programme would benefit students in terms of their academic progress. One facilitator said they could see academic progress within their school in terms of comparison with the previous year group:

> 'Well, in terms of our data for our reading we are progressing more than we had with last year's year seven cohort, so actually they are getting more from it in terms of being able to respond in their tests, their teacher assessments are looking really positive for their reading, their confidence—like particularly our lower—is coming along leaps and bounds.'

Some teachers did not feel that the programme would benefit their students' literacy skills because of the emphasis on team-working at the expense of time for individual activity like writing:

> 'I actually think that, in terms of literacy and building English skills, it's slightly lacking, because [there is] a lot of emphasis on teamwork, which is good, but a lot suffered as a result of the time when pupils got to sit quietly and write on their own.'

One positive consequence of the programme, which was mentioned by three of the teachers interviewed, was that students' social skills developed, potentially as a result of both the ability grouping and the teamwork aspects of the programme:

'It really encouraged the students to work with other students they wouldn't otherwise work with. I think that's very important with Year 7 because they get into cliques, don't they?'

'From what I can see, social skills-wise and working as a team, the sort of things that you can't record with data…. to me that has really worked, because I very rarely have an incident where I have a child saying, "I don't want to go in that group. I don't want to work with them". And that is usually what you get when you try and put children in groups.'

'I think they [Year 7s] will have formed relationships they might not otherwise have formed at secondary school. Yeah, I honestly did feel that and I think—they don't have as much after school as some other children so in lessons it's important for them to build social skills too.'

4.4 Pupils' experience of the programme

Pupils in both the intervention and control groups were asked, via the online survey, how much they enjoyed their English lessons. There were 398 respondents from the intervention group, and 779 respondents from the control group (see Appendix B for breakdown of number of responses per school). Overall, just over half (53%) of respondents were girls.

Pupils in the control group were more likely to say they enjoyed their English lessons a lot (25% vs 12%), with pupils in the intervention group more likely to say 'not much' (38% vs 25%).

There was no real difference between the groups on how difficult they found their English lessons, with the vast majority saying 'very easy' or 'quite easy' (76% intervention, 79% control).

Control group pupils were somewhat more positive about the books they read in English lessons, with 22% enjoying them 'very much' compared with 14% of pupils in the intervention group.

Pupils were asked to assess how much progress they had made over the year in different aspects of English. Whereas nearly all pupils in the control group thought they had progressed a lot or quite a lot in English overall (84%), this decreased to 72% of pupils in the intervention group. The difference was particularly marked for writing (with 37% of control pupils saying they had made 'a lot' of progress compared to only 22% for the intervention group), and for speaking (33% and 19% respectively).

When asked how much they agreed with various statements about their English lessons, intervention and control pupils had very similar opinions regarding the amount covered, the pace of lessons, and the number of tests. Intervention pupils were less likely to strongly agree that they did a lot of work in the lessons (15% intervention, 31% control) or learned a lot (26% and 38% respectively).

The responses from the students about the nature of their English lessons suggested that the pedagogy expected in Quest had been followed. Intervention pupils reported doing the following much more often than the control group: working in small groups/teams (50% versus 18% 'very often'), working with a partner (34% versus 16% 'very often'), and having discussions (62% versus 40% 'very often'). They worked in silence less often (11% versus 25% of the control group saying 'very often').

The majority of pupils liked working with partners or in small groups/teams 'very much' or 'quite a lot', regardless of which arm of the trial they were in. Although most students did not like working in silence, this was more evident for the intervention group (49% not liking it at all) than the control group (37%).

The control group were more likely to rate themselves 'very' good at English (26%) compared with the intervention group (9%), with the opposite pattern being seen for 'quite' good (58% control and 70% intervention).

Pupils in the intervention group were asked how many of their English lessons were Quest. Just under half (45%) said 'all' or 'most', with 27% saying 'some'. Five percent said none of their lessons were

Quest, and nearly a quarter (24%) did not know, which suggests either the Quest lessons were not very well flagged as such, or these students had not experienced them.

Almost half the students who said that some or most of their lessons were Quest found these lessons more enjoyable that their other English lessons. A slightly lower proportion (38%) said they were less enjoyable. The reasons given for finding Quest lessons more enjoyable most frequently mentioned them being more fun, particularly as regards working in groups:

> *'…because its more fun and everyone get a fair chance to say they [sic] ideas.*

> *'Because before we just did spellings and grammar when this is more fun.'*

> *'Some lessons were VERY [sic] fun where we had to work in teams and in some lessons we put that to the test.'*

> *'Because we always worked with different people every lesson also we did a lot of acting, group work, posters, serious work, funny work, watched films what was relivanted [sic] to the lessons and she was a nice teacher.'*

Others preferred non-Quest English lessons, complaining that it was boring (with particular mention of the books and the questions from the booklets) and that certain English skills (such as writing) were under-emphasised:

> *'It can get repeatative [sic] and boring at times with PEE ['Point, Evidence, Explanation'] paragraphs and random reporters.'*

> *'…reading books [in Quest] for little kids like 'The Iron Man' and boring books that don't make any sense like 'The Tempest'. We used to read story books for our actual age groups…*

> *'I prefer to do more writing. Id [sic] prefer to go back to normal English lessons.'*

> *'…because we were answering questions from the quest book over and over again which was quite boring.'*

There were mixed views about some of the Quest approaches:

> *'I don't like working in teams and in Quest lessons if you don't know the answer, they still pick on you and its annoying because you have to ask your group the answer.'*

> *'…because it's fun to talk about your opinions to other people and to share your idears [sic].'*

> *'Because you got to work in teams and get to know other peoples thoughts.'*

> *'Just that I don't learn as much when i'm reading a book then talking about what happened. Its just not helping us develop our grammar and spelling and how to use our punctuation. But when we do off topic lessons I learn about my spelling but quest does help our skill to read.'*

4.5 Control group activity

As the trial was designed with randomisation at the school level, the risk of contamination between groups was low, and there was no evidence that any schools in the control group were exposed to Quest.

Data was collected from teachers and students in the control group through the online survey. This allowed for comparison of standard practice in the control group with the intervention group. There were 11 responses from teachers in 5 intervention schools, and 35 responses from teachers in 10 control

schools. Teachers were asked how well their schools' current English scheme supported various skills and components (such as writing, vocabulary development, language acquisition, and independent work). Quest teachers rated their English scheme lower than control group teachers for writing and monitoring and assessment, but more strongly in terms of the following: reading fluency; speaking, talking and listening; language acquisition; collaborative work; and debate and discussion.

The effectiveness of the teaching in developing the various skills and components of English noted above is reflected in teacher feedback about how much time was devoted to these components. Teachers in the control group were spending more of their time on writing, monitoring and assessment—as well as independent work—than teachers in the intervention group. However, intervention teachers seemed to be covering more activities more frequently than control teachers, namely reading fluency; reading comprehension; vocabulary development; speaking and listening; language acquisition; collaborative work; and debate and discussion. The clearest differences concerned (a) writing (nearly all control teachers did this in 'every' or 'most' lessons, compared with about half for intervention teachers); (b) speaking and listening (over half the Quest schools did this every or most lessons compared with about one in three control teachers); (c) collaborative work (in every lesson for over half the Quest schools compared with less than a fifth of control schools); and (d) debate and discussion (in every or most lessons for 8/10 Quest schools, but only about a third of control schools).

When asked more general questions about their own teaching and classroom practice, Quest teachers seemed quite polarised when reporting whether Year 7 had shown more progress than expected, or whether they had enjoyed teaching Year 7 more than the previous year. They were less likely than control teachers to agree that their classroom management or confidence in teaching had increased during the year.

Conclusion

Key conclusions
1. The evaluation was unable to provide a secure estimate of the impact of Quest on reading comprehension outcomes among Year 7 pupils, primarily due to a high level of drop-out from the project.
2. Few, if any, schools implemented the programme as designed, suggesting that substantial adaptation may be required if the approach is to gain wider adoption in English schools.
3. Some of the main barriers to successful implementation included the difficulty in covering the expected material in a single school lesson; a perception of an insufficient focus on writing activities; and its adoption as a whole-year group intervention.
4. Many schools were positive about the range of resources provided by the programme, and about the co-operative learning aspects of the programme.
5. Though it is not possible to draw a conclusive statement about the impact of the programme, on average, pupils who received the programme made less progress than those who did not.

5.1 Limitations

The randomisation process resulted in intervention and control groups that were comparable at baseline in terms of pre-test (KS2) reading scores, however the control group had a higher proportion of girls, and a higher proportion of pupils eligible for free school meals.

The trial had a high level of drop-out, as well as differential attrition between the control and intervention groups (46% of schools in the intervention group, and 17% of schools in the control group dropped out after allocation). Differential measurement attrition was also present, with a larger proportion of missing data at post-test in the intervention schools compared to the control group (44% missing in the intervention group compared to 15% missing in the control group). While the reasons for this remain unclear, a plausible explanation may be that some schools only tested those ability groups who remained more engaged with the programme.

Given the extent of attrition and the differential responses from the control and intervention groups, the main effects reported have limited security. They were not statistically significant and, more importantly, given the level of attrition any estimates reported are likely to have been subject to some form of bias. This makes it impossible to draw any firm conclusions regarding the effectiveness of the Quest programme.

5.2 Interpretation

This report describes the implementation and impact of Quest during the first year of delivery in English secondary schools. Quest was based upon another programme (Reading Edge) that, according to a randomised trial of high poverty schools in the US, showed some evidence of effectiveness, including a moderate improvement on overall reading test scores. Quest's effectiveness as a reading intervention has not been demonstrated in the current evaluation, but this pilot programme, adapted for the English context, has identified significant implementation issues.

On average, pupils who received the programme scored lower in terms of reading comprehension compared to those who were taught English using standard practice. However, as there was a high level of attrition in the trial, it was not possible to conclude that the observed effect was actually due to the programme rather than through chance or some form of bias in the results. The programme did not appear to have differential effects according to pupil gender, free school meal eligibility, or prior attainment. Considerable caution is advised in interpreting these results, however, both in the light of the limitations discussed above and also given that overall implementation of the programme was poor. The current evaluation provides limited evidence towards establishing the effectiveness of the

programme as a reading intervention in English schools. Only one school was delivering daily Quest lessons, and a minority of the lessons observed were judged to be at the level of full implementation.

The process evaluation highlighted a number of issues that both undermined teacher confidence in the programme, and impacted upon the extent of implementation. The perceived weaknesses of Quest according to teachers were: that it was (a) overly prescriptive; (b) the lesson plans contained too much material to be covered in a standard 50 or 60 minute school period; and (c) it contained insufficient writing opportunities. A number of schools criticised the late arrival of texts and other resources, and the poor proof-reading of resources. Positive feedback was provided by the teachers in relation to the range of books and digital texts provided for pupils, although these could be improved for lower ability sets. On the whole, teachers were also positive about Quest*s* co-operative learning structure, particularly regarding teamwork and opportunities for discussion.

5.3 Future research and publications

The current evaluation adds limited evidence to our knowledge base of whole-year group approaches to reading interventions that involve co-operative learning, setting by ability, and concentrating most resources on those who are experiencing the greatest difficulties. The evaluation, however, highlights a number of important lessons for future evaluations. In terms of methodology, it highlights the importance of schools understanding the principle of intention-to-treat, particularly that all pupils are required to provide outcome data regardless of the extent of implementation of the programme. While online testing may be more efficient than traditional pen and paper methods, it is susceptible to technical difficulties and may be more prone to some form of bias through missing data and attrition. Ideally, if used, its use should be overseen by a member of the evaluation team.

In terms of the programme, Quest has only recently been adapted from the US to the English context, and the current efficacy trial was conducted during the first year of implementation. The current evaluation highlights the importance of careful pilot work prior to rigorous evaluation through a trial. This may be particularly salient when programmes with evidence of effectiveness are imported from other national contexts. Further work, taking into account feedback from teachers, is needed in order to adapt Quest to the English secondary school system. This includes reducing the amount of material to be covered in each lesson, and consideration of whether there is a need for more writing opportunities within the programme.

References

Chamberlain, A., Daniels, C., Madden, N.A. and Slavin, R.E. (2007) 'A randomised evaluation of the Success for All Middle School Reading Program', *Middle Grades Reading Journal*, 2(1), 1–22.

Cooper, J., Prescott, S., Cook, L., Smith, L. and Mueck, R. (1990) *Co-operative learning and college instruction*. California: California State University.

Hattie, J. (2009) *Visible Learning: A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.

Johnson, D.W., Maruyama, G., Johnson, R.T., Nelson, D. and Skon, L. (1981) 'Effects of co-operative, competitive and individualistic goal structures on achievement: A meta-analysis', *Psychological Bulletin*, 89, 47–62.

Kyndt, E., Raes, E., Lisamont, B., Timmers, F., Cascallar, E. and Dochy, F. (2013) 'A meta-analysis of the effects of face-to-face co-operative learning. Do recent studies falsify or verify earlier findings?', *Educational Research Review*, 10, 133–149.

Lou, Y., Abrami, P.C., Spence, J.C., Poulson, C., Chambers, B. and d'Apollonia, S. (1996) 'Within-class grouping: a meta-analysis', *Review of Educational Research*, 66, 423–458.

Qin, Z., Johnson, D.W. and Johnson, R.T. (1995) 'Co-operative versus competitive efforts and problem solving', *Review of Educational Research*, 65, 129–143.

Slavin, R.E. (1983) 'When does co-operative learning increase student achievement?', *Psychological Bulletin*, 94, 429–445.

Slavin, R.E., Chamberlain, A., Daniels, C., and Madden, N.A. (2009) 'The Reading Edge: A randomized evaluation of a middle school co-operative reading program',. *Effective Education,* 1 (1), 13–26.

Torgerson, C. J. and Torgerson, D. J. (2007) 'The use of minimization to form comparison groups in educational research', *Educational Studies*, 33 (3), 333–337.

Thanh, P.T.H., Gillies, R. and Renshaw P. (2008) 'Co-operative learning and academic achievement of Asian students: A true story', *International Education Studies*, 1, 83–88.

Appendix 1: Parent information and consent form

INSTITUTE FOR EFFECTIVE EDUCATION, THE UNIVERSITY OF YORK

Quest

Information Sheet for Parents/Guardians

Dear Parent/Guardian,

We would like to request your permission for your child to take part in an educational research study. The following information explains why the research is being done and what it would involve for your child.

The Institute for Effective Education (IEE) is part of the University of York. It aims to find out what works in teaching and learning and why, and then use the evidence to improve education.

This study is being done to assess the effectiveness of Quest, a Year 7 English programme designed to help pupils quickly improve literacy skills. Students are grouped according to their reading skills and use co-operative learning. We are conducting this study with Year 7 classes in 25 secondary schools in England. The headteacher of your child's school has agreed to participate in this study.

Schools have been randomly assigned either to use the Quest programme starting in September 2013 and ending in June 2014, or to be a comparison school, teaching English and literacy as usual. In June 2014, teachers will administer short literacy tests to all of the Year 7 pupils in the participating schools. After this all participating schools, including the comparison schools, will be able to use the Quest programme if they wish.

As part of the research, your child will take a short literacy test in June 2014. You may choose not to permit your child to participate in the testing, but they will participate in the literacy programme (unless the school is in the comparison group), as this will be part of the taught curriculum throughout Year 7. Your child's scores will be seen only by those who mark the assessments and by your child's teachers. Then pupils' names will be replaced with code numbers and no individual pupil's data will appear in any report about the research study.

If you do not want your child to participate, please complete and sign the attached opt-out form by [X pm on Y date]. A pupil's right to withdraw will be respected.

Please could you tell your child about the research study – that it aims to evaluate the effectiveness of a literacy programme and will involve a short test at the end of the school year - and explain that they have the right to withdraw from the study at any time.

If you have a concern or question about your child's participation in this study, please contact Pam Hanley (e-mail: pam.hanley@york.ac.uk Tel:01904 328165) or Emma Marsden, the head of the Education Ethics Committee (email: emma.marsden@york.ac.uk) about the study.

INSTITUTE FOR EFFECTIVE EDUCATION

Parent/Guardian opt-out form

If you do not **permit your child to participate in the study, please complete this form and return it to your child's teacher by 3:00 pm on** (one week after receipt).

I **do not** wish my child to take part in the research project. (If you do not want your child to take part, they will complete another piece of work set by their teacher when the other pupils are doing the assessment)

Pupil's name: ....
(Please print clearly)

Form teacher's Name: ....

Parent's/Guardian's name: ....
(Please print clearly)

Parent's/Guardian's signature: ....

Date………………………………………………………………………………………………

Appendix 2: Full fidelity ratings

SCHOOL (ability)	1 (high)	1 (low/ mid)	2 (mid)	2 (low)	3 (high)	3 (low/m id)	7 (low)	16 (high)	19 (high)	21 (high)	21 (lo w)	23 (hig h)	23 (lo w)
Lesson moves at an appropriate pace	3	3	3	1	2	2	2	3	3	3	3	3	2
Teacher has students read books at students' level	na	na	na	na	3	1	3	3	3	na	3	3	na
Teacher circulates and supports students	3	3	3	2	3	3	3	3	3	3	3	3	3
Teacher has good classroom manageme nt	2	3	2	2	2	2	2	3	3	3	3	3	3
Teacher models skills e.g. using Think Alouds	na	na	na	1	0	na	1	na	0	3	3	1	2
Pupils work in heterogene ous partners/te ams	2	2	0	0	2	3	3	3	3	3	3	1	3
Pupils display co-operative behaviour during group work	3	1	1	0	1	1	2	2	2	3	na	3	1
Teachers use Think-Pair-Share	na	0	0	na	2	1	3	na	0	2	na	0	2
Teachers use Whole Group Response	na	0	3	na	na	3	na	3	1	0	3	3	0
Teachers use Random Reporter	na	0	0	na	2	3	3	2	1	0	3	3	3
Pupils receive team points	1	3	0	0	0	0	3	2	1	0	3	0	1
Pupils are engaged during teacher presentatio ns	2	2	3	na	3	3	1	2	3	3	3	3	1
Pupils are familiar with routines	3	3	1	na	2	2	3	3	3	3	3	3	3
Teachers recognise the learning of team members, not just co-operative behaviour	3	2	0	1	0	2	3	2	3	3	3	1	1

Overall implementation rating	3	2	1	1	1	1	2	3	2	2	3	2	2

Appendix 3: Student survey responses across schools

School	Intervention status	Number of student responses	% of total responses
1	Intervention	17	1.4%
2	Intervention (withdrawn)	37	3.1%
3	Intervention	56	4.8%
4	Control	0	-
5	Control	93	7.9%
6	Intervention (withdrawn)	19	1.6%
7	Intervention	105	8.9%
8	Control	0	-
9	Control	0	-
10	Control	69	5.9%
11	Intervention (withdrawn)	63	5.4%
12	Control	0	-
13	Control	152	12.9%
14	Intervention (withdrawn)	1	0.1%
15	Control (withdrawn)	130	11.0%
16	Intervention	1	0.1%
17	Control	64	5.4%
18	Intervention (withdrawn)	0	-
19	Intervention	0	-
20	Control	70	5.9%
21	Intervention	10	0.8%
22	Control	69	5.9%
23	Intervention (withdrawn)	2	0.2%
24	Control	160	13.6%
25	Intervention (withdrawn)	59	5.0%

Appendix 4: Security classification of trial findings

1.Design: What is the quality of the design of the evaluation?	2. Power: What is minimum detectable effect at the start?	3. Attrition: What is the level of drop out from the evaluation?	4. Balance: Adjustment to rating to account for balance?	5. Threats to validity: Adjustment to account for issues with interpretation?

Evaluation design — Implementation — Analysis and interpretation

Rating	1. Design	2. Power (MDES)	3. Attrition	4. Balance	5. Threats to validity
5 🔒	Fair and clear experimental design (RCT)	< 0.2	< 10%	Well-balanced on observables	No threats to validity
4 🔒	Fair and clear experimental design (RCT, RDD)	< 0.3	< 20%		
3 🔒	Well-matched comparison (quasi-experiment)	< 0.4	< 30%		
2 🔒	Matched comparison (quasi-experiment)	< 0.5	< 40%		
1 🔒	Comparison group with poor or no matching	< 0.6	< 50%		
0 🔒	No comparator	> 0.6	> 50%	Imbalanced on observables	Significant threats

The final security rating for this trial is 1 🔒. This means that the conclusions have low security.

The trial was designed as an efficacy trial and could achieve a maximum of 5 🔒. The trial was moderately underpowered because fewer schools were recruited than intended for the MDES of 0.23 (25 schools instead of 30). There was substantial attrition of schools, with 26% of schools dropping out. The pupils were imbalanced at baseline as well as in the analysis sample—in terms of the proportion eligible for FSM and gender—resulting in the loss of a further padlock. A final padlock was lost because of Treat to Validity: substantial numbers of pupils were missing post-test data (42% overall), and this was differential by arm (intervention 56% missing post-test, control 28% missing post-test). Therefore, the overall padlock rating is 1 🔒.

Appendix 5: Cost rating

Cost ratings are based on the approximate cost per pupil per year of implementing the intervention over three years. More information about the EEF's approach to cost evaluation can be found on the EEF website. Cost ratings are awarded as follows:

Cost rating	Description
£	*Very low:* less than £80 per pupil per year.
£ £	*Low:* up to about £200 per pupil per year.
£ £ £	*Moderate:* up to about £700 per pupil per year.
£ £ £ £	*High:* up to £1,200 per pupil per year.
£ £ £ £ £	*Very high:* over £1,200 per pupil per year.