Education Endowment Foundation

Grammar for Writing

Evaluation Report and Executive Summary

February 2014

Independent evaluators:

The University of York

Durham University

Professor David Torgerson

Professor Carole Torgerson

Co-authors: Dr Natasha Mitchell, Ms Hannah Buckley, Ms Hannah Ainsworth, Ms Clare Heaps, Dr Laura Jefferson

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.

We aim 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 founded in 2011 by lead charity The Sutton Trust, in partnership with Impetus Trust (now part of Impetus – The Private Equity Foundation), with a £125m grant from the Department for Education. With investment and fundraising income, the EEF intends to award as much as £200m by 2026. Together, the EEF and Sutton Trust are the Government-designated What Works Centre for Improving Education Outcomes for School-Aged Children.

Literacy Catch-up

In May 2012 the Department for Education awarded the EEF a further £10 million for a grants round dedicated to literacy catch-up projects for children at the primary-secondary transition. The projects funded within this round aimed to identify effective ways to support pupils who do not achieve Level 4 in English by the end of Key Stage 2.

For more information please contact:

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

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

About the evaluators

David Torgerson

Professor David Torgerson, Director of the York Trials Unit, joined the Centre for Health Economics at the University of York in 1995 and became the Director of the York Trials Unit in 2002. Originally a health economist, he is a trial methodologist and has published widely on the design and conduct of randomised controlled trials including the book *Designing Randomised Controlled Trials in Health, Education and the Social Sciences* (Palgrave Macmillan, 2008). David has over 200 peer-reviewed publications.

Contact details:

David Torgerson
Professor, Director of York Trials Unit
York Trials Unit
Dept of Health Sciences
University of York
YO10 5DD

p: 01904 321340
e: david.torgerson@york.ac.uk

Carol Torgerson

After taking degrees in English Literature and Modern English Poetry from the Universities of York and Stirling, Carole Torgerson taught English in a wide variety of schools in England and Scotland. She undertook her doctorate in design issues in experimental research at the University of Sheffield before starting her academic career. She was Reader in Evidence-based Education at the University of York, before moving to The University of Birmingham as Professor of Experimental Design and then to Durham University where she is Professor of Education.

Contact details:

Carole Torgerson
Professor of Education
School of Education
Durham University
Leazes Road
DH1 1TA

p: 01901 334 8382
e: carole.torgerson@durham.ac.uk

Contents

Executive Summary

The project

Grammar for Writing is a literacy intervention that aims to improve writing skills of Year 6 pupils by providing contextualised grammar teaching. It encourages pupils to improve how their writing communicates with the reader by making connections between a linguistic feature and the effect it has on writing, rather than by focusing on grammatical inaccuracies. It can be delivered as both a universal and targeted intervention.

In this evaluation, the programme was trialled as both a whole class and small group approach, with pupils expected to achieve between Level 3 and 4b at Key Stage 2. The intervention involved 15 guided writing sessions, delivered over four weeks, in June/July 2013, and involved 53 primary schools from four geographical regions across England. Teachers in intervention classes received three days of training in the Grammar for Writing approach.

What impact did it have?

The overall effect size of the class-level intervention compared to the 'business as usual' control was small (+0.10, estimated two additional months' progress) and statistically insignificant (i.e., may have occurred by chance). A larger effect was observed when children were taught in small groups (+0.24, estimated three additional months' progress), although this is very similar to the difference in effect size between the small group intervention and the whole class approach. This suggests that the observed gains in writing outcomes are likely to be as a result of teaching pupils in small groups, rather than any intrinsic benefit of teaching Grammar for Writing. There is, therefore, little evidence that the intervention provided additional gains in writing outcomes.

The study does demonstrate that small group teaching is an effective strategy to increase writing skills among pupils between Level 3c and 4b, which is reasonably consistent with the emerging evidence base for small group tuition.

The subgroup analysis showed no evidence of the intervention differentially benefiting pupils eligible for free school meals.

Group	N	Effect size	Estimated months' progress	95% confidence interval (CI)*	Evidence Strength**
Class level intervention vs. control	1,982	+0.10	+2	-0.09 to 0.30	★★★☆☆
Small group intervention vs. control	817	+0.24	+3	0.00 to 0.49	★★★☆☆
Small group intervention vs. whole group intervention	631	+0.21	+3	0.04 to 0.38	★★★☆☆
Class level intervention vs. control without small group	1,772	+0.06	+1	-0.15 to 0.28	★★★☆☆

*Effect sizes with confidence intervals that pass through 0 are not 'statistically significant', which means the difference could be due to chance.

Evidence ratings are a new measure under development based on a number of factors including study type, size and drop-out. Ratings are provisional and are not given for sub-group analyses, which will always be less secure than overall findings. For more information about ratings visit: **www.educationendowmentfoundation.org.uk/evaluation.

How secure is this finding?

This evaluation was set up as an effectiveness trial to test the impact of a four-week version of Grammar for Writing, delivered with the developer leading the recruitment and retention of the schools and participants and the training, and overseeing the provision of the intervention. EEF effectiveness trials aim to test whether an intervention can work at scale, in real-world conditions.

In each school, one Year 6 class was randomly assigned to the intervention group and one Year 6 class was randomly assigned to continue with teaching as usual. Within the intervention class, eligible pupils were randomised on an individual basis to receive either whole group Grammar for Writing teaching or the whole group approach plus small group Grammar for Writing teaching. At the end of the intervention period all pupils were asked to complete the Progress in English 11 (Long Form) Test developed by GL Assessment, as a measure of general writing ability.

This study was a well conducted, relatively large study with independent randomisation. Intention to treat analysis was used (i.e. pupils were compared in the groups to which they were originally randomly assigned); blinded marking of test papers was undertaken, and correct statistical adjustment was made for class randomisation. There was about 20% drop-out in terms of the primary outcome, but this was evenly spread between the groups and there was no evidence that it introduced selection bias. The process evaluation indicates the programme was delivered with good fidelity (i.e. as intended by the developer). Overall, this indicates the findings are secure.

A previous trial evaluation of Grammar for Writing showed a statistically significant impact among older children, but methodological issues in the way the data were analysed means that there are doubts about the validity of the findings.

How much does it cost?

The cost of the approach is estimated at £20 per pupil. This estimate includes resources (estimated at 50p per pupil), direct salary costs of teachers (£250), initial training (£700) and on-going monitoring and support (£250). Estimates are based on a school delivering the intervention to 60 pupils, with six out-of-class intervention pupils and two teachers trained.

Key Conclusions
1. Grammar for Writing is not effective in improving general writing among Year 6 pupils when delivered as a whole class intervention over four weeks. 2. Grammar for Writing is modestly effective in improving writing as a small group intervention, although this is likely to be a result of small group teaching, rather than an intrinsic benefit of Grammar for Writing itself.
3. The evidence for Grammar for Writing from this evaluation is insufficient to recommend widespread adoption among Year 6 pupils.
4. Small group teaching amongst Levels 3c to 4b pupils does seem an effective strategy to increase writing skills.
5. Possible further research question: Is small group teaching for children at Levels 3c to 4b effective and cost-effective over a longer time period?

Introduction

Intervention

The Grammar for Writing intervention is a curriculum intervention aimed at improving writing skills by providing contextualised grammar teaching.

Background evidence

The final year at primary school (Year 6) and the move to secondary school is commonly known as 'transition'. The transition starts early in Year 6 and does not end until sometime after pupils have settled into their new school (Evangelou et al., 2008). Recent figures suggest that Key Stage 2 (KS2) attainment in English has remained static over the last few years: 79% of pupils achieved Level 4 or above in 2005, with a slight rise to 82% in 2011 (Ofsted, 2012). In terms of writing standards, there is some evidence of a decrease overall, and English tends to be weaker amongst pupils who are eligible for free school meals (FSM) (Ofsted, 2012).

In 2012, the Education Endowment Foundation (EEF) funded the University of York and Durham University to independently evaluate the Grammar for Writing intervention as delivered by the intervention developers based at the University of Exeter.

Grammar for Writing is a curriculum intervention aimed at improving writing skills by providing contextualised grammar teaching. The intervention in this trial was a modified version of an existing grammar intervention aimed at improving writing skills in older children; this modified version was aimed at children in Year 6 who were less able writers. An evaluation of the existing intervention was undertaken by the developers (Jones et al., 2012; Myhill et al., 2012) who found some evidence that the intervention was effective in enhancing writing performance in Year 8 pupils. Myhill et al. (2012) also found the intervention benefited sub-groups differently: it appeared to assist able writers more than weaker writers. However the previous trial did not use intention to treat analysis (it removed a school allocated to the control group due to 'poor' implementation) and it did not adjust for the clustered nature of the data (statistical analysis assumed individual randomisation, when cluster randomisation had been used so this would have produced biased standard errors).

The current study was developed in the light of this evidence. Unlike the previous trial, the current trial focuses on a younger year group, and adds to the evidence-base around interventions to improve grammar in writing in this age group.

The implementation team (Exeter) were responsible for developing and delivering the Grammar for Writing intervention.

Evaluation objectives

The objective of the independent evaluation was to test the effectiveness of the intervention. The research question was:

What is the effectiveness of the small group Grammar for Writing intervention, when compared with a 'business as usual' control group, on writing skills of participating children?

Project team

The evaluation team was responsible for the design, conduct, analysis and reporting of the independent evaluation.

Chief Investigators

Professor David Torgerson (DT), Director, York Trials Unit, Department of Health Sciences, University of York, Heslington, York, YO10 5DD. T: 01904 321340 E: david.torgerson@york.ac.uk

Professor Carole Torgerson (CT), School of Education, Durham University, Leazes Road, Durham, DH1 1TA. T: 0191 3348382 E: carole.torgerson@durham.ac.uk

Trial Manager

Natasha Mitchell (NM), Research Fellow, York Trials Unit, Department of Health Sciences, University of York, Heslington, York, YO10 5DD. T: 01904 321655 E: natasha.mitchell@york.ac.uk

Trial Statistician

Hannah Buckley (HB), Trainee Statistician, York Trials Unit, Department of Health Sciences, University of York, Heslington, York, YO10 5DD. T: 01904 321512 E: hannah.buckley@york.ac.uk

Researchers

Hannah Ainsworth (HA), Research Fellow, York Trials Unit, Department of Health Sciences, University of York, Heslington, York, YO10 5DD. T: 01904 328158 E: hannah.ainsworth@york.ac.uk

Laura Jefferson (LJ), Research Fellow, York Trials Unit, Department of Health Sciences, University of York, Heslington, York, YO10 5DD. T: 01904 321511 E: laura.jefferson@york.ac.uk

Clare Heaps (CH), Research Associate, School of Education, Durham University, Leazes Road, Durham, DH1 1TA. T: 0191 3348332 E: C.N.Heaps@dur.ac.uk

Implementation Team

The implementation team was responsible for school recruitment and on-going relationships with schools, informing parents and pupils, intervention development (including writing a detailed description of the intervention to allow others, if necessary, to be able to replicate the intervention in other areas), intervention training and delivery, and baseline data collection.

Professor Debra Myhill, Professor of Education, Associate Dean for Research, SSIS, Graduate School of Education, University of Exeter, Heavitree Road, Exeter EX1 2LU. T: 01392 724767 E: D.A.Myhill@exeter.ac.uk

Dr Susan Jones, Senior Lecturer in Education, SSIS, Graduate School of Education, University of Exeter, Heavitree Road, Exeter EX1 2LU. T: 01392 724767 E: susan.m.jones@ex.ac.uk

Dr Ros Fisher, Associate Professor of Education, SSIS, Graduate School of Education, University of Exeter, Heavitree Road, Exeter EX1 2LU. T: 01392 724767 E: r.j.fisher@ex.ac.uk

Dr Anthony Wilson, Lecturer in Education, SSIS, Graduate School of Education, University of Exeter, Heavitree Road, Exeter EX1 2LU. T: 01392 724767 E: a.c.wilson@ex.ac.uk

Lucy Oliver, Research Assistant, SSIS, Graduate School of Education, University of Exeter, Heavitree Road, Exeter EX1 2LU. T: 01392 724767 E: ljo204@ex.ac.uk

Helen Lines, Research Assistant, SSIS, Graduate School of Education, University of Exeter, Heavitree Road, Exeter EX1 2LU. T: 01392 724767 E: h.e.lines@ex.ac.uk

National Association of Teachers of English (NATE)

NATE was a project partner and assisted the University of Exeter with the recruitment and subsequent liaison with schools. They also facilitated communication of the project outcomes.

Methodology

Trial Design

A pragmatic cluster-randomised trial with a split-plot design was employed. Recruitment targeted schools with two Year 6 classes. The classes were randomised into two groups: a Grammar for Writing group and a 'business as usual' control group; one class within each school was randomised to each group. Within the intervention classes, individual children who met the inclusion criteria were randomised to receive either the whole class form of the intervention alone or to receive the whole class intervention plus a small group intervention. This design is known as a 'partial split plot': it is a variant of a factorial design due to its combination of cluster and individual randomisation. A cluster-randomised design was required in this case as the intervention is class based, which precluded the use of individual randomisation. However, the addition of the individual level randomisation allowed for further investigation into the effect of Grammar for Writing when delivered as a small group intervention. The design, therefore, allowed us to examine the class-level effects of Grammar for Writing by comparing the intervention classes with the control classes. Additionally, it allowed us two further comparisons of interest. First, it meant that we could disentangle any 'small group' treatment effects by comparing the small group pupils in the intervention group with their peers, who were not in small groups, in the control class. Second, we could also ascertain whether there was any additional advantage of delivering Grammar for Writing in a small group compared with the whole class: this was assessed by comparing those in the intervention classes who were randomised to small group with those randomised to whole class.

The trial was designed, conducted and reported to CONSORT standards (Altman et al., 2011) in order to minimise all potential threats to internal validity, such as selection bias and a range of post randomisation biases (Cook and Campbell, 1969; Shadish, Cook and Campbell, 2002; Torgerson and Torgerson, 2008). In this way, unbiased estimates of the impact of the intervention are provided.

Ethical Review

The trial protocol and information sheets were reviewed by two ethics bodies: the Durham University School of Education Ethics Committee and the Chair of Department of Health Sciences (University of York) Research Ethics and Governance Committee.

Recruitment

The developer of the intervention led on recruitment and retention of schools and participants. The evaluation team (University of York and Durham University) and the implementation team (University of Exeter), in collaboration with NATE, jointly provided information documentation about the trial to schools (Appendix 1). Schools who wanted to take part were asked to sign an 'Agreement to Participate Form' (Appendix 2) to ensure they agreed to all trial-related procedures. The schools targeted for this trial were those with (a) a high proportion of pupils eligible for free school meals, (b) a high proportion of children achieving Level 3 or borderline Level 4 in English and (c) ideally two Year 6 classes.

Participating primary schools informed parents of all pupils in Year 6 about the study using material provided by the evaluation team and the University of Exeter (Appendix 3). Parents had the opportunity to withdraw their child's data from being used in the evaluation (opt out) prior to randomisation. Participating primary schools then shared pupil data with the evaluation team (including pupil name, unique pupil number (UPN), date of birth (DOB), free school meals status (FSM), Key Stage 2 (KS2) English teacher assessment from Dec 2012).

Eligibility

School inclusion criteria: Primary schools were eligible to take part in the trial if they agreed to all trial procedures, including informing parents, provision of pupil data, randomisation and implementation of the intervention as allocated.

Pupil inclusion criteria: Within the intervention class, pupils were eligible for individual randomisation if they were expected to achieve Level 3c, Level 3b, Level 3a, Level 4c or Level 4b in English by the end of Key Stage 2 (based on teacher assessment).

School exclusion criteria: Primary schools were excluded from participating in the trial if they did not agree to all points listed in the 'Agreement to Participate Form' or if they were not able to carry out testing at the end of the intervention period.

Pupil exclusion criteria: Pupils were excluded from individual randomisation if they were expected to achieve below Level 3 or above Level 4b. Exclusion also occurred if parents/guardians returned an opt-out form to the school, and in these instances no data were provided to the evaluators. Those predicted to achieve below Level 3 were excluded from testing as it was thought the post-testing could have caused undue anxiety.

Interventions

The intervention was designed by the implementation team from the University of Exeter. It involved the use of teaching materials with embedded grammar teaching, with the aim of improving writing. The implementation team developed 15 sequential guided writing sessions, and the embedded grammar aspects encouraged pupils to make connections between a linguistic feature and the effect it has in writing (Jones et al., 2012). The intervention focused on encouraging pupils to actively make grammatical choices which would affect how their writing would communicate to the reader; it did not focus on pupil's grammatical errors or any inaccuracies (Jones et al., 2012). Year 6 classes randomised to the intervention used their literacy class time to deliver the intervention. Eligible pupils within the intervention class were randomised to the small group with the intention that this group would receive the intervention in the whole class setting as well as additional intervention in a small group. However, we cannot be certain that this small group teaching occurred in all schools as there was no fidelity assessment in terms of adherence to the trial design. Pupils randomised to the 'business as usual' group received their usual literacy lesson as planned by their teacher.

Outcomes

The Progress in English (PiE) test (PiE 11: Second Edition Long Form (LF) Test, GL Assessment), was the main test used to determine literacy outcome. The test includes both narrative and non-narrative exercises and assesses both reading and writing skills including areas such as spelling, grammar and comprehension. The Progress in English test was the only test available to the evaluation team (in order to comply with EEF testing policy) which included a writing component. Tests were marked by GL Assessment blind to allocation (i.e. markers did not know whether test papers were from either the intervention or control pupils).

Primary outcomes

The primary outcome was extended writing score which refers to the combined raw score on the two extended writing tasks (Exercises 5 and 6) from the PiE 11 LF. Exercise 5 has a total possible 20 marks and involves writing a persuasive letter; the maximum in Exercise 6 is 12 marks and assesses informative writing. Overall, the extended writing task score can be in the range 0 to 32, with a higher score representing higher attainment.

Secondary outcome

Reading score, the combined raw score on the reading tasks (Exercises 3, 3x, 4 and 4x) was used as the secondary outcome. Exercise 3 (comprising Exercises 3 and 3x) has a total possible 19 marks and assesses reading comprehension of a narrative. Exercise 4 (comprising Exercises 4 and 4x) has a total possible 13 marks and assesses non-narrative reading comprehension. Overall, reading score can range between 0 and 32, with a higher score representing better attainment.

Spelling and grammar score, the combined raw score on the spelling and grammar tasks (Exercises 1 and 2) was chosen as a further secondary outcome. Exercise 1 has a total possible 10 marks and assesses spelling; Exercise 2 has a total possible 10 marks and assesses grammar. This means the spelling and grammar score combined can range from 0 to 20, with a higher score representing higher attainment.

Delivery of outcomes

Teachers were asked to deliver the outcome tests. They were not blind to the group allocation of the children. However, they were asked to deliver the test under ‘exam’ conditions with the whole class sitting the test at the same time.

Sample size

The focus of this trial was on pupils who were performing between Level 3c and Level 4b, therefore the sample size calculation was based on this subgroup of children.

For the purposes of calculating the sample size it was assumed 60 schools would be recruited with an average of 54 pupils per school; this would result in a total sample size of 3,240 pupils. Assuming 27 pupils per class and an intra-cluster correlation coefficient of 0.19 would lead to a design effect of 5.94. When divided into the total sample size, this produces an *effective sample size* of 546 pupils. However, assuming a pre- and post-test correlation of 0.70, the effective sample size increases to 1,070. We allowed for an attrition rate of 10% meaning the final effective sample size was 964 pupils. This would allow a difference of 0.18 standard deviations to be detected in the writing scores of the intervention and control classes, should one exist.

For the individually randomised component of the trial, it was assumed that there would be approximately 8 children per class in the 60 classes (480 pupils in total) and that there would be a pre- and post-test correlation of 0.70; this would increase the effective sample size from 480 to 942. We allowed for an attrition rate of 10% which gave an effective sample size of 848, meaning that a difference of 0.20 of a standard deviation (80% power; 2p = 0.05) in writing scores could be detected between the two randomised groups, if such a difference existed. If there were a modest intra-cluster correlation of 0.05 remaining, despite individual randomisation, then the effective sample size might decline to 630 participants as there would be a design effect of 1.35. This effective sample size would allow for detection of an effect size of 0.23 standard deviations (80% power, 2p = 0.05), should one exist.

Randomisation

Randomisation was conducted at two levels: class and individual. At the class level, one class was randomised to receive the intervention and one class continued with 'business as usual' within each school. This randomisation was conducted using stratification by school with a fixed block size of 2. Further randomisation within the intervention class was conducted at the individual level for pupils predicted to achieve between Level 3c and Level 4b in KS2 Writing. Eligible pupils were assigned to either receive the whole class form of the intervention only, or to receive the whole class intervention plus small group intervention through deterministic minimisation within schools. Minimisation is a technique that ensures balance between the groups by using an arithmetical algorithm. The algorithm calculates the balance on specified variables after each individual has been allocated such that the next allocated individual minimises any chance imbalance between the groups (Torgerson & Torgerson, 2008). Gender and predicted KS2 writing level were used as minimisation factors with two and three levels respectively. As each small group needed to contain between 4 and 6 pupils and due to the fact that class size varied, different allocation ratios were used depending on the number of eligible pupils in the intervention class at each school. In total, five allocation ratios were employed as below:

Number eligible pupils	Allocation ratio (Whole: Small)
6 or less	1:2
7-11	1:1
12-18	2:1
19-24	3:1
25-30	4:1

Both the class- and individual-level random assignments were conducted by the trial statistician (HB) based at York Trials Unit. Class randomisation was conducted in Stata® version 12 (Stata Corporation, College Station, Texas, USA); individual-level minimisation was conducted using minimPy (http://sourceforge.net/projects/minimpy/). The class-level allocation occurred first; individual-level allocation occurred after the trial statistician received pupil baseline data. At each stage of randomisation the evaluation team provided this information to the implementation team for them to disseminate this information to the schools.

Analysis

Analysis was conducted in Stata® version 13 (Stata Corporation, College Station, Texas, USA) using the principles of intention to treat, meaning that all classes and pupils were analysed in the group to which they were randomised irrespective of whether or not they actually received the intervention and irrespective of implementation fidelity. Statistical significance was assessed at the 5% level, unless otherwise stated.

Effect sizes were calculated and are presented alongside 95% confidence intervals. Effect size is defined as:

$$\Delta = \frac{\beta_{\text{intervention}}}{\sigma_{\varepsilon}}$$

where $\beta_{\text{intervention}}$ is the difference in mean score between the intervention and control groups and σ_{ε} is the residual standard deviation. The residual standard deviation was used rather than the more usual pre-test standard deviation as there was no equivalent pre-test value. Numerical values used to calculate the effect sizes for each analysis can be found in Appendix 4.

The test and outcomes were examined for ceiling or floor effects using summary statistics and graphical representations. Intra-cluster correlation coefficients (ICCs) were estimated and are presented alongside 95% confidence intervals.

Cluster-level analysis

Primary analysis

The primary objective of this part of the trial was to investigate the effectiveness of the Grammar for Writing on the writing skills of all pupils at Level 3 and above. The difference in writing scores between pupils in the intervention classes and those in the 'business as usual' classes was compared using a multilevel regression analysis to allow for the hierarchical nature of the data. The model used extended writing score as the response variable, with group allocation, gender, FSM status, English as an additional language (EAL) status, month of birth and predicted KS2 score included as fixed effects; school and class were included as random effects.

Secondary analyses

The primary analysis was undertaken a total of four times. The first analysis used reading score as the response and the second spelling and grammar score in order to assess the impact of the intervention in terms of the secondary outcomes. The third analysis compared pupils allocated to receive additional small group teaching of the intervention with those in the control group at Levels 3c, 3b, 3a, 4c or 4b;. The effect of the intervention in terms of extended writing score was also analysed in the sub-group of pupils who were eligible for FSM through the inclusion of an interaction term in a final iteration of the primary analysis.

Individual-level analysis

Primary analysis

The primary objective of this trial was to investigate the effectiveness of the small group form of the intervention on the writing skills of eligible pupils. The difference in writing scores between pupils allocated to the whole class plus small group intervention and those to whole class intervention only was compared using a multilevel regression analysis with extended writing score as the response variable. Group allocation, gender, FSM status, EAL status, month of birth and predicted KS2 score were used as fixed effects in the model, with class as a random effect. Although the trial was randomised at the individual level, because children were taught in classes or small groups there would still be a clustering of outcomes, hence the need to use multilevel regression methods.

Secondary analyses

The primary analysis was undertaken a total of four times. The first analysis used reading score as the response and the second spelling and grammar score in order to assess the impact of the intervention in terms of the secondary outcomes. The third analysis compared pupils allocated to receive additional small group teaching of the intervention with those in the control group at levels 3c, 3b, 3a, 4c or 4b. The effect of the intervention in terms of extended writing score was also analysed in the sub-group of pupils who were eligible for FSM through the inclusion of an interaction term in a final iteration of the primary analysis.

Process evaluation methods

Fidelity was assessed for every intervention class in the trial using a measure devised by the implementation team. No fidelity assessment was conducted in the control arm. The measure consists of three component scores relating to (a) the use of grammar terms, (b) linking grammar effects in writing and (c) using talk to develop discussion about choices and effects. Each of these components was rated between 1 and 3, with 1 corresponding to ‘rarely’, 2 corresponding to ‘partially as planned’ and 3 corresponding to ‘as planned’. As such, the fidelity score could range between 0 and 9, with higher scores corresponding to higher fidelity.

Results

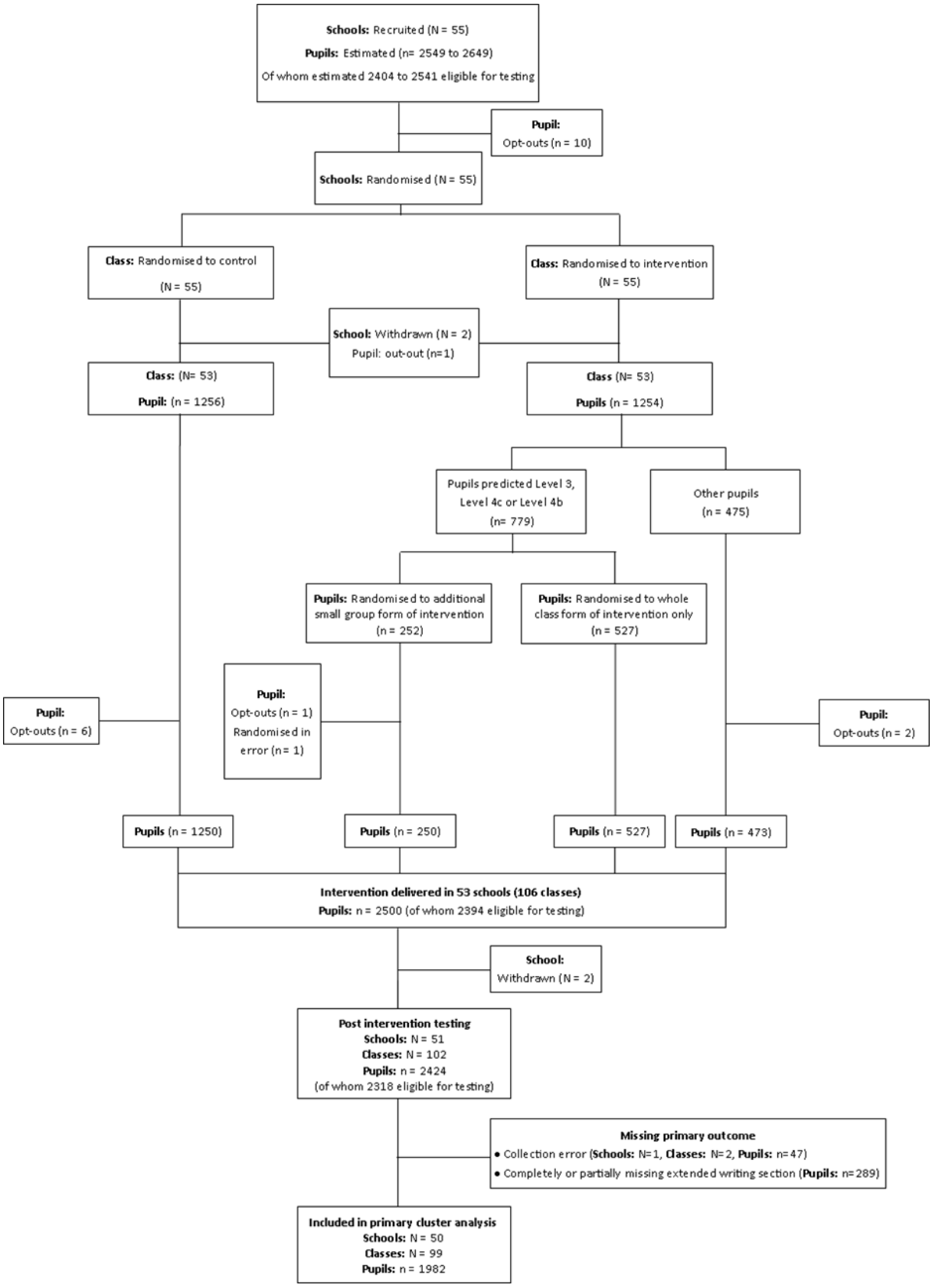
Recruitment and follow-up of participants

The implementation team, in collaboration with NATE, recruited appropriate schools and pupils. School recruitment took place between January and March 2013. Four geographical areas were targeted: Sheffield, London, West Midlands and the South West. Originally it was proposed to recruit 60 schools with two classes per school. However, due to time constraints it was only possible to recruit 55 schools each with two classes. A total of four schools (eight classes) withdrew from the trial: two withdrawals occurred post cluster-level randomisation and two occurred post individual-level randomisation (during or after the intervention delivery period). This left 51 schools and 102 classes involved in the trial at the point of testing (Figure 1). At the time of individual level randomisation, 2,510 pupils were included in the trial. At the start of intervention delivery, 2,500 pupils were involved in the trial (of whom 2,394 were eligible for testing). By the testing period, 2,424 pupils remained, and of these 2,318 were eligible for testing (i.e., predicted to achieve Level 3 or above).

For the individual level part of the trial 779 pupils were randomised, all of whom were eligible for testing. Of these, 31 pupils were from one of the two schools which withdrew post individual randomisation: one opted-out and another was randomised in error. Extended writing score was missing for 115 of the remaining 746 pupils leading to an attrition rate of 19.0%.

As the two schools which withdrew post cluster randomisation did not provide information on the number of pupils who would have been in the trial, it is not possible to say with certainty how many pupils were involved at the point of recruitment. Using the fact that, for schools within the sample, the number of pupils per school ranged between 14 and 64, it is estimated that there many have been between 2,549 and 2,649 pupils at the time of recruitment. Using the fact that between 1 and 9 pupils were below Level 3 for the other schools in the trial and taking into account the 21 opt-outs it is estimated that between 2,404 and 2,541 of these pupils were eligible for testing. This leads to an individual pupil attrition rate at the cluster level which could range between 17.6% and 22.0%.

Figure 1. CONSORT flow diagram



Baseline characteristics

Table 1 shows characteristics of the school which were recruited to the trial. There was a large proportion of missing data (a minimum of 47.3% missing) for each variable. The mean school size was around 439 pupils (SD 158.37). Around a third of pupils in the recruited schools were eligible for FSM and approximately 45% were of minority ethnic origin. Both of these percentages are considerably higher than the national averages in January 2013 which were reported at 19.2% and 28.5% respectively (Department for Education, 2013).

Table 1: Characteristics of recruited schools

	School
	N = 55
Number of pupils on roll	
Mean (SD)	438.8 (158.37)
Med (Min, Max)	424 (213, 947)
Missing (%)	26 (47.3)
Percentage of pupils eligible for FSM	
Mean (SD)	33.2 (17.93)
Med (Min, Max)	30 (12, 93)
Missing (%)	28 (50.9)
Percentage of pupils from a minority ethnic group	
Mean (SD)	44.6 (36.38)
Med (Min, Max)	42.6 (2, 98)
Missing (%)	30 (54.5)
Percentage of pupils supported by School Action Plus (without statement of SEN)	
Mean (SD)	12.6 (10.45)
Med (Min, Max)	9.9 (5, 55)
Missing (%)	31 (56.4)

	School
Special measures	
Yes (%)	0 (0.0)
No (%)	25 (45.5)
Missing (%)	30 (54.5)

The average class size in both the intervention and control classes was 23.6 pupils (Table 2). The average number of pupils predicted to achieve between Level 3c and Level 4b was also consistent between the control and intervention classes at between 14 and 15 pupils. Of the 53 schools for which pupil data were provided, classes at 32 schools involved pupils who were predicted to achieve below Level 3. Not all schools included classes with pupils predicted below Level 3 as some organised in mixed ability groups and others in literacy groups. It was estimated that around 8 of the participating schools organised in literacy groups and that 44 of the schools organised in mixed ability groups (data for one school was missing).

Table 2: Baseline characteristics of participating schools and classes

	School	Intervention class	Control class
	N=53	**N=53**	**N=53**
Pupils in trial			
Mean (SD)	47.2 (11.78)	23.6 (5.78)	23.6 (6.37)
Med (Min, Max)	50 (14, 64)	24 (8, 33)	26 (6, 31)
Pupils between Level 3c and Level 4b			
Mean (SD)	29.0 (9.90)	14.7 (5.58)	14.4 (5.71)
Med (Min, Max)	28 (7, 54)	14 (2, 27)	14 (4, 27)

Baseline summary statistics and pupil characteristics (Table 3) are based on the 2,394 pupils at Level 3 or above from the 53 schools for whom baseline data were provided; no data were available relating to two schools which withdrew from the trial following the cluster-level randomisation but before individual randomisation. Table 3 shows baseline characteristics by cluster-level allocation (i.e. intervention and control) both as randomised and as analysed in the primary cluster-level analysis. In relation to the demographic characteristics – FSM status, Pupil Premium (PP) status, EAL status, month of birth and predicted KS2 writing level at baseline – proportions of pupils within each category were similar between the intervention and control arms, both as randomised and as analysed in the primary cluster-level analysis.

Table 3: Baseline pupil level characteristics

	As randomised (All Level 3 or above) Frequency (%)		As analysed (primary cluster analysis) Frequency (%)	
	Intervention	**Control**	**Intervention**	**Control**
	n =1,194	**n** = 1,200	**n** = 1,004	**n** = 978
Gender				
Male	609 (51.0)	617 (51.4)	507 (49.5)	500 (51.1)
Female	585 (49.0)	583 (48.6)	497 (50.5)	478 (48.9)
Missing	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
FSM				
Eligible	420 (35.2)	392 (32.7)	328 (32.7)	295 (30.2)
Not eligible	774 (64.8)	808 (67.3)	676 (67.3)	683 (69.8)
Missing	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Pupil premium				
Eligible	446 (37.4)	425 (35.4)	348 (34.7)	324 (33.1)
Not eligible	748 (62.7)	775 (64.6)	656 (65.3)	654 (66.9)
Missing	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
English as an additional language				
EAL	494 (41.4)	516 (43.0)	408 (40.6)	423 (43.3)
Non-EAL	700 (58.6)	684 (57.0)	596 (59.4)	555 (56.8)
Missing	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Predicted KS2 writing level				
Level 3c	48 (4.0)	65 (5.4)	41 (4.1)	45 (4.6)
Level 3b	78 (6.5)	80 (6.7)	62 (6.2)	62 (6.3)
Level 3a	128 (10.7)	113 (9.4)	93 (9.3)	88 (9.0)
Level 4c	278 (23.3)	256 (21.3)	228 (22.7)	203 (20.8)

	As randomised (All Level 3 or above) Frequency (%)		As analysed (primary cluster analysis) Frequency (%)	
Level 4b	245 (20.5)	248 (20.7)	207 (20.6)	209 (21.4)
Level 4a	189 (15.8)	185 (15.4)	168 (16.7)	165 (16.9)
Level 5 or above	228 (19.1)	253 (21.1)	205 (20.4)	206 (21.1)
Missing	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Month of birth				
Sep – Nov	331 (27.7)	315 (26.3)	281 (28.0)	260 (26.6)
Dec – Feb	305 (25.5)	299 (24.9)	260 (25.9)	230 (23.5)
Mar – May	278 (23.3)	293 (24.4)	234 (23.3)	245 (25.1)
Jun – Aug	280 (23.5)	293 (24.4)	229 (22.8)	243 (24.8)
Missing	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)

Of the 2,394 pupils eligible for inclusion in the primary analysis, 412 (17.2%) were missing the primary outcome of extended writing score. Reasons and distribution of missing data are presented in Table 4.

Table 4: Reasons for missing data relating to the cluster analysis

Reason for missing outcome	Frequency	% of missing outcomes (412)	% of those eligible for analysis (2,394)
Collection error	47	11.4	2.0
School withdrawal	76	18.4	3.2
Did not attempt any of the extended writing questions (inc. potential absence)	176	42.7	7.4
Only partially completed the extended writing questions	113	27.4	4.7
Total	**412**		

Table 5 shows baseline characteristics by individual-level allocation (i.e. whole class intervention only, and whole class intervention plus small group form of intervention) both as randomised and as analysed in the primary individual-level analysis. These summary statistics are based on the 777 pupils from 53 intervention classes who were predicted to achieve between Level 3c and Level 4b in KS2 writing and for whom baseline data were provided. In relation to the demographic characteristics (FSM status, PP status, EAL status, month of birth and predicted KS2 writing level at baseline),

proportions of pupils within each category were similar between those allocated to remain in the whole group and those randomised to receive the additional form of the intervention. This is the case both as randomised and as analysed in the primary individual-level analysis.

Table 5: Baseline characteristics of pupils who were individually randomised

	As randomised Frequency (%)		As analysed (primary individual-level analysis) Frequency (%)	
	Small	**Whole**	**Small**	**Whole**
	n = 250	**n** = 527	**n** = 210	**n** = 421
Gender				
Male	144 (57.6)	295 (56.0)	121 (57.6)	233 (55.3)
Female	106 (42.4)	232 (44.0)	89 (42.4)	188 (44.7)
Missing	0 (0.0)	(0.0)	0 (0.0)	0 (0.0)
FSM				
Eligible	98 (39.2)	225 (42.7)	74 (35.2)	171 (40.6)
Not eligible	152 (60.8)	302 (57.3)	136 (64.8)	250 (59.4)
Missing	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Pupil Premium				
Eligible	105 (42.0)	233 (44.2)	80 (38.1)	175 (41.6)
Not eligible	145 (58.0)	294 (55.8)	130 (61.9)	246 (58.4)
Missing	0 (0.0)	0 (0.0)	(0.0)	0 (0.0)
English as an additional language				
EAL	108 (43.2)	232 (44.0)	89 (42.4)	185 (43.9)
Non-EAL	142 (56.8)	295 (56.0)	121 (57.6)	236 (56.1)
Missing	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Predicted KS2 writing level				
Level 3c	18 (7.2)	30 (5.7)	17 (8.1)	24 (5.7)
Level 3b	19 (7.6)	59 (11.2)	16 (7.6)	46 (10.9)

	As randomised Frequency (%)		As analysed (primary individual-level analysis) Frequency (%)	
Level 3a	44 (17.6)	84 (15.9)	31 (14.8)	62 (14.7)
Level 4c	86 (34.4)	192 (36.4)	72 (34.3)	156 (37.1)
Level 4b	83 (33.2)	162 (30.7)	74 (35.2)	133 (31.6)
Level 4a	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Level 5 or above	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Missing	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Month of birth				
Sep – Nov	57 (22.8)	126 (23.9)	50 (23.8)	100 (23.8)
Dec – Feb	60 (24.0)	132 (25.0)	48 (22.9)	107 (25.4)
Mar – May	61 (24.4)	120 (22.8)	53 (25.2)	98 (23.3)
Jun – Aug	72 (28.8)	149 (28.3)	59 (28.1)	116 (27.6)
Missing	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)

Of the 777 pupils eligible for inclusion in the primary analysis, 146 (18.8%) were missing the primary outcome of extended writing score. Reasons and distribution of missing data are presented in Table 6.

Table 6: Reasons for missing data relating to the individual analysis

Reason for missing outcome	Frequency	% of missing outcomes (146)	% of those eligible for analysis (777)
Collection error	11	7.5	1.4
School withdrawal	31	21.2	4.0
Did not attempt any of extended writing questions (inc. potential absence)	66	45.2	8.5
Only partially completed extended writing questions	38	26.0	4.9
Total	**146**		

Table 7 presents summary statistics of data relating to the teachers which were collected after the CPD days. As classes were randomised (as opposed to the teachers), intervention or control refers

here to the class the teacher taught in practice. Teacher data were collected from 53 of the 55 schools which were involved in the cluster-level randomisation. At one school, given the re-randomisation of pupils into three classes (further details provided in the fidelity section below), data are recorded for two control teachers. This is also the case for another school where staff job-share in the control class. In one further school, two teachers co-taught the intervention class and data were provided for both.

The proportion of male teachers in the intervention arm was slightly higher than that in the control arm at 20.4% compared with 16.4%. An assessment of grammar knowledge was conducted by the implementation team using a test with a score ranging from 0 to 30, with higher marks relating to a higher level of grammar knowledge. Data relating to this test are missing for 15 of the 109 teachers (13.8%). The mean grammar knowledge score was similar between arms although a higher proportion of intervention teachers had less than 5 years teaching experience (40.7% compared with 16.4%). The distribution of teacher age was fairly similar between the intervention and control groups; however, there were more intervention teachers aged between 26 and 30 years than control teachers (35.2% compared with 20.0%) and more control teachers aged between 36 and 40 than intervention teachers (20.3% compared with 3.6%). Missing data were more common in relation to control teachers, with over a quarter of data missing on each variable.

Table 7: Characteristics of teachers by allocated groups

	Intervention	Control
	n = 54	**n =** 55
Gender		
Male (%)	11 (20.4)	9 (16.4)
Female (%)	43 (79.6)	28 (50.9)
Missing (%)	0 (0.0)	18 (32.7)
Grammar knowledge score		
Mean (SD)	20.0 (3.89)	19.0 (4.37)
Med (Min, Max)	21 (10, 29)	20 (6, 27)
Missing (%)	0 (0.0)	15 (27.3)
Teaching experience (years)		
0 – 5 (%)	22 (40.7)	9 (16.4)
6 – 10 (%)	15 (27.8)	19 (34.5)
11 – 15 (%)	7 (13.0)	5 (9.0)
16 – 20 (%)	5 (9.3)	3 (5.5)
21 – 25 (%)	1 (1.9)	4 (7.3)
26 – 30 (%)	1 (1.9)	1 (1.8)

	Intervention	Control
31 – 35 (%)	3 (5.6)	0 (0.0)
Missing (%)	0 (0.0)	14 (25.5)
Age of teacher (years)		
21 – 25 (%)	5 (9.3)	3 (5.5)
26 – 30 (%)	19 (35.2)	11 (20.0)
31 – 35 (%)	8 (14.8)	12 (21.8)
36 – 40 (%)	11 (20.4)	2 (3.6)
41 – 45 (%)	6 (11.1)	8 (14.5)
46 – 50 (%)	1 (1.9)	2 (3.6)
51 – 55 (%)	4 (7.4)	2 (3.6)
56 – 60 (%)	0 (0.0)	0 (0.0)
61 – 65 (%)	0 (0.0)	1 (1.8)
Missing (%)	0 (0.0)	14 (25.5)

Outcomes and analysis

The test and outcomes were assessed for ceiling or floor effects using histograms and summary statistics. No evidence of either such an effect was found (details in Appendix 5).

Due to a collection error, post-test data were available for 50 of the 51 schools which remained in the trial at the point of testing. Pupils who were predicted to achieve below Level 3 in KS2 writing at baseline are excluded from all analyses.

Based on the results from these 50 schools, intra-cluster correlation coefficients (ICCs) were estimated (Table 8). These were somewhat larger than the one that was used for the sample size calculation estimates (i.e. 0.19). The correlation between outcome and the predicted KS2 level was also lower than expected (Spearman's Rho 0.54).

Table 8: Estimated intra-cluster correlation coefficients (ICCs)

	n	School ICC (95% CI)	Class ICC (95% CI)
Total raw score	1,977	0.21 (0.13 to 0.28)	0.27 (0.21 to 0.34)

	n	School ICC (95% CI)	Class ICC (95% CI)
Primary outcome (extended writing score)	2,033	0.26 (0.17 to 0.34)	0.32 (0.25 to 0.39)

Cluster-level analysis

Raw, unadjusted mean post-test scores are presented in Table 9 by trial arm. Scores were similar in both allocated groups at 22.8 out of 32 marks (SD 4.85) in the intervention group and 22.6 marks (SD 4.88) in the control group. Proportions of those completing the extended writing questions were similar in both allocated arms. Mean reading score and mean spelling and grammar score were also similar between the arms, as were proportions completing.

Table 9: Unadjusted average scores for the intervention and control groups

	Intervention	Control	Overall
Pupils predicted Level 3 and above	**n = 1,194**	**n =1,200**	**n = 2,394**
Primary outcome			
Extended writing score	**n** = 1,004	**n** = 978	**n** = 1,982
Mean (SD)	22.8 (4.85)	22.6 (4.88)	22.7 (4.86)
Med (Min, Max)	23 (7, 32)	23 (9, 32)	23 (7, 32)
Completely missing (%)	147 (12.3)	152 (12.7)	299 (12.5)
Partially missing (%)	43 (3.6)	70 (5.8)	113 (4.7)
Complete (%)	1,004 (84.1)	978 (81.5)	1,982 (82.8)
Secondary outcomes			
Reading score	**n** = 867	**n** = 847	**n** = 1714
Mean (SD)	18.4 (5.18)	18.2 (5.22)	18.3 (5.2)
Med (Min, Max)	19 (4, 31)	18 (4, 30)	18 (4, 31)

	Intervention	Control	Overall
Completely missing (%)	135 (11.3)	141 (11.8)	276 (11.5)
Partially missing (%)	192 (16.1)	212 (17.7)	404 (16.9)
Complete (%)	867 (72.6)	847 (70.6)	1714 (71.6)
Spelling and grammar score	**n** = 1,025	**n** = 1,051	**n** = 2,076
Mean (SD)	11.4 (5.08)	11.5 (4.92)	11.4 (5.00)
Med (Min, Max)	12 (0, 20)	12 (0, 20)	12 (0, 20)
Completely missing (%)	140 (11.7)	140 (87.6)	280 (11.7)
Partially missing (%)	29 (2.4)	9 (0.8)	38 (1.6)
Complete (%)	1,025 (85.9)	1,051 (87.6)	2,076 (86.7)

Primary analysis

The primary analysis adjusted for baseline-predicted KS2 writing level, gender, FSM status, EAL status and month of birth. After exclusion for missing data relating to any of these variables or the response, analysis was conducted on 1,982 pupils from 99 classes; 1,004 from the intervention group and 978 from the control group. There was little evidence of a difference in extended writing score between the allocated groups, with a non-significant increase of 0.34 marks for those in the intervention group when compared with those in the control group (p=0.30, 95% CI: -0.30 to 0.98). This relates to an effect size of 0.10 (95% CI: -0.09 to 0.30).

Secondary analyses

Reading

After exclusion for missing data, analysis was conducted on 1,714 pupils from 99 classes (and 50 schools): 867 from the intervention group and 847 from the control group. There was little evidence of a difference in reading score between the allocated groups, with a non-significant increase of 0.38 marks for those in the intervention group when compared with those in the control group (p=0.14, 95% CI: -0.12 to 0.88). This relates to an effect size of 0.10 (95% CI: -0.03 to 0.24).

Spelling and grammar

After exclusion for missing data, analysis was conducted on 2,076 pupils from 100 classes; 1,025 from the intervention group and 1,051 from the control group. There was little evidence of a difference in

spelling and grammar score between the allocated groups, with a non-significant increase of 0.04 marks for those in the intervention group when compared with those in the control group (p=0.88, 95% CI: -0.44 to 0.51). This relates to an effect size of 0.01 (95% CI: -0.14 to 0.16).

Control versus small group form of intervention

Pupils in intervention classes who were randomised to receive the small group form of the intervention, and pupils in the control class predicted to achieve between Level 3c and Level 4b in KS2 writing, were eligible for inclusion in this analysis. Exclusion occurred if a pupil had missing data relating to any included covariates or relating to the response (extended writing score). In total, 1,012 pupils were eligible for inclusion in this analysis: 250 from the small group intervention group and 762 from the control group. After exclusions, data from 817 pupils from 99 clusters: 210 from the small group intervention and 607 from the control group, were used to fit the model. There was some evidence of a statistically significant difference in extended writing score between the allocated groups, with a statistically significant increase of 0.78 marks for those who received the small group form of the intervention when compared with those in the control group (p=0.05, 95% CI: -0.01 to 1.56). This relates to an effect size of 0.24 (95% CI: 0.00 to 0.49).

Subgroup analysis

A pre-specified subgroup analysis was conducted. The effect of the intervention on pupils eligible for FSM was assessed through the inclusion of an interaction term in a repetition of the primary analysis. Statistical significance was assessed at the 10% level. Despite there being no evidence of an intervention effect, there was some evidence of a statistically significant interaction between allocated group and FSM status (p=0.08), suggesting the intervention had a different effect on FSM and non-FSM pupils. Table 10 shows the marginal mean extended writing scores for those receiving FSM and those not receiving FSM, by trial arm: the scores are higher for non-FSM pupils than for those eligible for FSM. There was no evidence of a significant effect in any of the groups.

Table 10: Marginal mean and effect sizes relating to extended writing scores for FSM and non-FSM pupils

	Intervention	Control
Eligible for FSM		
Marginal mean	21.8 (95% CI: 21.0 to 22.6)	21.9 (95% CI: 21.1 to 22.7)
Effect size	0.02 (95% CI: -0.21 to 0.24)	0.03 (95% CI: -0.11 to 0.18)
Not eligible for FSM		
Marginal mean	23.2 (95% CI: 22.4 to 23.9)	22.7 (95% CI: 21.9 to 23.4)
Effect size	0.16 (95% CI: -0.05 to 0.37)	Reference group

Individual-level Analysis

Raw, unadjusted mean post-test scores are presented in Table 11 by individual level allocation for those in the intervention class. Mean extended writing score was similar in both allocated groups at 21.7 out of 32 marks (SD 4.36) for those allocated to the small group form of the intervention and 20.9

marks (SD 4.33) for those randomised to receive the whole group form of the intervention only. Proportions of those completing the extended writing sections of the post-test were similar between arms. Mean reading score and mean spelling and grammar score were similar between the arms as were proportions of those completing.

Table 11: Average writing scores comparing small group versus whole class

	Small group	Whole class	Overall
	n = 250	**n =527**	**n = 777**
Primary outcome			
Extended writing score	**n** = 210	**n** = 421	**n** = 631
Mean (SD)	21.7 (4.36)	20.9 (4.30)	21.1 (4.33)
Med (Min, Max)	22 (10, 32)	21 (7, 32)	21 (7, 32)
Completely missing (%)	33 (13.2)	75 (14.2)	108 (13.9)
Partially missing (%)	7 (2.8)	31 (5.9)	38 (4.9)
Complete (%)	210 (84.0)	421 (79.9)	631 (81.2)
Secondary outcomes			
Reading score	**n** = 167	**n** = 336	**n** = 503
Mean (SD)	16.2 (5.49)	16.2 (4.44)	16.2 (4.51)
Med (Min, Max)	16 (6, 29)	16 (4, 30)	16 (4, 30)
Completely missing (%)	30 (12.0)	66 (12.5)	96 (12.4)
Partially missing (%)	53 (21.2)	125 (23.7)	178 (22.9)
Complete (%)	167 (66.8)	336 (63.8)	503 (64.7)
Spelling and grammar score	**n** = 210	**n** = 449	**n** = 659
Mean (SD)	9.7 (5.08)	9.2 (4.78)	9.3 (4.67)
Med (Min, Max)	10 (1, 19)	9 (0, 20)	9 (0, 20)
Completely missing (%)	33 (13.2)	67 (12.7)	100 (12.8)

	Small group	Whole class	Overall
Partially missing (%)	7 (2.8)	11 (2.1)	18 (2.3)
Complete (%)	210 (84.0)	449 (85.2)	659 (84.8)

Primary analysis

The primary analysis adjusted for baseline-predicted KS2 writing level, gender, FSM status, EAL status and month of birth. After exclusion for missing data relating to any of these variables or the response, analysis was conducted on 631 pupils from 50 schools (and 50 classes); 210 allocated to additional small group intervention and 421 from the whole group only arm. There was some evidence of a difference in extended writing score between the allocated groups, with a statistically significant increase of 0.67 marks for those in the small group arm when compared with those in the whole group arm (p=0.02, 95% CI: 0.12 to 1.23). This relates to an effect size of 0.21 (95% CI: 0.04 to 0.38).

Secondary analyses

Reading

After exclusion for missing data, analysis was conducted on 503 pupils from 49 schools (and 49 classes); 167 allocated to additional small group intervention and 336 from the whole group only arm. There was no evidence of a difference in reading score between the allocated groups, with a non-significant increase of -0.03 marks (i.e. a decrease of 0.03 marks) for those in the intervention group when compared with those in the small group arm when compared with those in the whole group arm (p=0.94, 95% CI: -0.75 to 0.69). This relates to an effect size of -0.01 (95% CI: -0.20 to 0.18).

Spelling and Grammar

After exclusion for missing data, the analysis was conducted on 659 pupils from 49 schools (and 49 classes): 210 allocated to additional small group intervention and 449 from the whole group only arm. There was little evidence of a difference in spelling and grammar score between the allocated groups, with a non-significant increase of 0.48 marks for those in the intervention group when compared with those in the small group arm when compared with those in the whole group arm (p=0.11, 95% CI: -0.10 to 1.07). This relates to an effect size of 0.14 (95% CI: -0.03 to 0.31).

Subgroup analysis

A pre-specified subgroup analysis was conducted. The effect of the intervention on pupils eligible for FSM was assessed through the inclusion of an interaction term in a repetition of the primary analysis. Statistical significance was assessed at the 10% level. There was no evidence of a statistically significant interaction between allocated group and FSM status (p=0.54) suggesting the intervention did not have a differential effect dependent on FSM status.

Primary analysis with exclusion of intervention small group

The intention to treat cluster analysis demonstrated an effect size of 0.10. However, because there is a 'small group effect' potentially driving this non-significant effect size we repeated the analysis removing pupils who had been randomised to have the small group intervention. The analysis was conducted on 1,772 pupils; 978 from the control group and 794 from the intervention group. There was no evidence of a difference in extended writing score between the allocated groups, with a non-significant increase of 0.20 marks for those in the intervention group compared with those in the control group (p=0.57, 95% CI: -0.48 to 0.87). This relates to an effect size of 0.06 (95% CI: -0.15 to 0.28).

Fidelity

Fidelity scores were available for 52 of the 55 randomised schools. Two scores were missing due to school withdrawal before the start of the intervention, and the third due to withdrawal of a school during the intervention and before fidelity assessment. A fidelity score was available for the fourth school which withdrew from the trial due to the timing of the fidelity assessment. The minimum fidelity score recorded was 4 out of 9. The maximum and most frequently recorded score was 9, with 56.4% of schools being judged to have delivered as planned in relation to all three components. The mean fidelity score was 8.2 (SD 1.27) and the median score was 9.

One school requested three teaching groups after both levels of randomisation had occurred. The school allowed the evaluation team to create the new teaching groups randomly, but this meant that pupils potentially did not receive the condition to which they were originally assigned at both the individual and cluster level. At one school data were provided on incorrect classes. This was only discovered after cluster randomisation and after the first CPD day. The randomisation of the correct classes to intervention or control resulted in the previous intervention teacher teaching the control class and vice versa; hence the allocations were switched in practice. This meant that eligible pupils in the control class needed to be individually randomised for practical reasons. One school did not have enough pupils eligible for individual randomisation in the intervention class to teach a small group due to pupil extraction hence all pupils received the intervention at the class level.

Conclusions and implications

We have undertaken a large pragmatic randomised controlled trial of Grammar for Writing in Year 6 pupils. Our data suggest only a relatively small effect size (approximately 0.10 standard deviations difference) on the GL Assessment measure between the classes randomised to receive the intervention and those continuing with 'business as usual'. This difference was not statistically significant, with a 95% confidence interval ranging from -0.09 to 0.30, suggesting that this difference may have occurred due to chance. Indeed, when the small group children were excluded from the intervention group, the effect size was reduced to 0.06 of a standard deviation difference.

When children were taught in small groups there was a larger effect size of between 0.21 to 0.25, which did not materially differ in the comparisons between small groups versus large intervention groups, or small groups versus large control groups. This suggests, therefore, that the difference found in the small group Grammar for Writing is as a consequence of teaching children in small groups *per se* rather than any intrinsic benefit of teaching grammar in small groups.

Although we found little evidence that Grammar for Writing was effective as measured by the GL Assessment outcome in Year 6 pupils, we did find that teaching children in small groups of about 4-6 children per group did suggest an improvement in writing skills by around a quarter of a standard deviation compared with similar children taught in class sizes of approximately 25. This finding supports previous evidence that small group teaching is effective (EEF, 2013), although this benefit needs to be set against the increased cost of teaching children in small groups. However, there remains an alternative explanation for the impact of small groups. Whilst some schools delivered small group teaching within the same time allocation for literacy, other schools may have delivered additional teaching. Consequently, the apparent benefit of small group teaching may be due to additional teaching and not entirely due to being taught in small groups.

Strengths

In the design and conduct of our study we used best practice as defined by the CONSORT guidelines for randomised controlled trials. Importantly, we used independent concealed allocation to ensure that the schools and children were allocated without the possibility of bias. We used the principles of intention to treat by including all consenting children and schools in the final analysis. We pre-specified our main outcome and wrote a statistical analysis plan before we observed the data. We also used an independent company to mark the test papers, blind to the allocated group.

Limitations

Although our trial was relatively large, with over 100 classes and more than 2,400 pupils, it was not possible to recruit to the target of 120 classes in 60 schools. Furthermore, the actual intra-cluster correlation coefficients (ICCs) were somewhat larger than our predicted ICCs, which would have reduced our statistical power. In terms of attrition, we lost four schools after randomisation. Two schools withdrew from the study after cluster-level randomisation, two after individual-level randomisation, and for a fifth school the post-test data were not retrieved by the testing company. We also lost a number of pupils, for the main outcome, who did not complete all of the relevant questions on the post-test, and so were excluded from the analysis. However, we do not think that these post-randomisation exclusions are likely to have introduced bias as there is no reason to suppose that their loss was linked to the intervention. Reassurance for lack of selection bias due to attrition is given in

the baseline tables where there is little difference between the analysed groups: this suggests that the effect of attrition on observed variables was similar between the two groups, implying that selection bias was not a factor.

Although the test papers were marked blindly, they were delivered to the children by teachers who were not blind to group allocation. To reduce the possibility of teacher bias we gave instructions that the children sat the tests under 'exam' conditions. However, we cannot exclude the possibility that teachers may have given inappropriate help to some children whilst sitting the test.

The design meant that both the intervention and control classes were nested within the same school. Consequently we cannot completely exclude the possibility that the relatively small effect size difference between the groups may be as a consequence of contamination between the intervention and control teachers. If this did occur, however, it would suggest the transmission of the intervention between intervention and control teachers was as effective as the dedicated training sessions that intervention teachers attended. Furthermore, because the effect size after removing the small group effect was so slight (0.06), a significant proportion of control teachers must have been contaminated.

Generalisability of results

A wide range of schools across England were recruited, consequently our findings should be applicable to most English primary schools, particularly those in inner-city urban areas or schools with a high proportion of pupils belonging to minority ethnic groups or eligible for FSM.

Further research

Our study did not find sufficient evidence to support the use of Grammar for Writing with Year 6 pupils. We did find some evidence, however, which suggests that small group teaching may have modest effects among Year 6 pupils who were between Levels 3c and 4b in the Standard Assessment Tests (SATs). In our study, such children only had approximately one term's exposure to being taught in small groups. It might be useful to look at, say, a full year's exposure to small group teaching for these children and to estimate the effectiveness and cost-effectiveness of such an approach.

Conclusion

In conclusion, we found a small (effect size 0.10) impact of Grammar for Writing in our intention to treat analysis, which was not a statistically significantly difference and is, in part, explained by the small group impact of a subsample of children. Small group teaching may have had a modest benefit and would merit further study.

References

Altman, D.G., Moher, D. and Schulz, K.F. (2012), 'Improving the reporting of randomised trials: the CONSORT Statement and beyond', Statist. Med., 31: 2985–2997.

Cook, T.D. and Campbell, D. (1969), *Quasi-experimentation: Design and Analysis Issues for Field Settings*, Boston: Houghton Mifflin.

Department of Education (2013), 'Schools, Pupils and their Characteristics: January 2013', (Reference IdL SFR21/2013), Statistical First Release.

Education Endowment Foundation (EEF) (2013), 'Small group tuition, The Sutton-Trust-EEF Teaching and Learning Toolkit'.

Evangelou, M.., Taggart, B., Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I. (2008), 'What Makes a Successful Transition from Primary to Secondary School?: Findings from the Effective Pre-school, Primary and Secondary Education 3-14 (EPPSE) Project', Department for Children, Schools and Families.

Jones, S., Myhill, D.A., Bailey, T. (2012), 'Grammar for writing? An investigation of the effects of contextualised grammar teaching on students' writing', Reading and Writing, DOI 10.1007/s11145-012-9416-1.

Myhill, D.A., Jones, S.M., Lines H & Watson A. (2012), 'Rethinking grammar: the impact of embedded grammar teaching on students' writing and students' metalinguistic understanding' Research Papers in Education, 27:2, 139-166.

Ofsted (2012), 'Moving English forward: action to raise standards in English', (110118).

Schagen, I.A. (2004), *But What Does It Mean? The Use of Effect Sizes in Educational Research*, Slough: NFER.

Shadish W.R., Cook, T.D. & Campbell, D.T. (2002), 'Experimental and Quasi-Experimental Designs for Generalized Causal Inference', Boston, MA: Houghton, Mifflin and Company.

Torgerson, D.J. and Torgerson, C.J. (2008), *Designing Randomised Trials in Health, Education and the Social Sciences: An Introduction*, Basingstoke and New York: Palgrave Macmillan.

Torgerson, C.J., Wiggins, A., Torgerson, D.T., Ainsworth, H., Barmby, P., Hewitt, C., Jones, K., Hendry, V., Askew, M., Bland, M., Coe, R. Higgins, S., Hodgen, J., Hulme, C. and Tymms, P. (2011), 'The Every Child Counts Independent Evaluation Report', Department of Education.

Appendix 1: Project Briefing Sheet for Schools

CHOICE AND CONTROL: CONTEXTUALISING GRAMMAR WITHIN WRITING

Centre for Research in Writing

Recently, our research team completed a large national study which investigated whether drawing attention to specific grammar features during the teaching of writing, might help raise student attainment in writing. The results showed a significant improvement for all students, with a higher rate for able students; it also revealed the influence of the teacher's own grammatical knowledge on students' learning.

In partnership with the National Association for the Teaching of English (NATE), we have now been awarded funding by the *Education Endowment Foundation* to undertake another project to explore whether this approach will work for writers who are both socially disadvantaged and currently under-achieving in writing. In this project, we will be working with year 6 classes, that have a high proportion of students who receive Free School Meals and are achieving Level 3 or borderline Level 4 in writing. We will be designing Guided Writing interventions for use in the classroom with this group and out-of-class interventions offering additional support: these interventions will address the particular writing needs of these writers. Part of the intervention will be providing professional support for the teachers involved with three training days, and one day where each individual teacher is mentored and given feedback on an observation lesson. This will include specific subject knowledge development on the grammar used in the intervention. We hope that the way this project is designed with CPD on grammar and writing will itself be a significant advantage for teachers involved and will benefit all children: we cannot be sure, of course, whether the intervention will work for these most disadvantaged and weak writers.

What kind of schools are we looking for?

Because this project focuses on students who are below or borderline level 4 in writing and eligible for Free School Meals, we are looking for schools who have relatively high numbers of students in this category, and, ideally, who have two year 6 classes, one which will be an intervention class, the other a comparison class.

What commitment would this project require?

Enthusiasm for the project and for your own professional learning

Attendance at the three CPD training days and the plenary conference

Willingness to participate in coaching and teaching of a 'trial' lesson

Willingness to implement the Guided Writing and out-of-class interventions during June 2012

Willingness to allow the random allocation of one Year 6 class to the intervention and one year 6 class to be a comparison group

Willingness to allow random allocation of the Level 3 FSM writers in one class to the intervention

Willingness to allow the administration of a writing test by external consultant in early July 2012

Provision of baseline data about your class

When will this project take place?

The project begins in January but the three CPD days will be in March, April and May. The intervention will be implemented after the summer half term.

Is there funding to support my involvement?

Yes, there is supply cover to cover your attendance at the CPD training days plus another 9 days of supply cover to support the running of the intervention. You might, for example, need to bring in a supply teacher to allow you to do the one-to-one interventions. We will also provide the results to the school for the children in the study.

Evaluation

There will be an independent evaluation of the project conducted by the University of York and Durham University. The Universities are supported by the Educational Endowment Foundation to undertake the evaluation but have played no role in the development or implementation of the intervention. The evaluation team, led by Professor Carole Torgerson (Durham University) and Professor David Torgerson (University of York) will provide further information of on the design and implementation of the evaluation. Pupils' test responses and any other pupil data will be treated with the strictest confidence. The responses will be sent to and marked by GL Assessment and accessed by Durham University. Named data will be matched with the National Pupil Database and shared with Durham University and EEF. No individual school or pupil will be identified in any report arising from the research

If you would like to know more, please contact Debra Myhill (d.a.myhill@ex.ac.uk) or Susan Jones (susan.m.jones@ex.ac.uk)

Appendix 2: Primary School Agreement to Participate Form

Evaluation of University of Exeter's *Grammar for Writing* intervention

Primary School Agreement to Participate

☐ I confirm that I have read and understood the information sheet for the above evaluation and have had the opportunity to ask questions;

☐ I understand that all children's results will be kept confidential and protected using encryption software and that no material which could identify individual children or the school will be used in any reports of this evaluation;

☐ I agree to provide baseline data about pupils in Year 6 to the evaluation team, University of Exeter and EEF (excluding any pupils for whom opt out forms have been returned);

☐ I understand that named baseline data will be matched with the National Pupil Database and shared between the evaluation team, University of Exeter and EEF;

☐ I agree to random allocation to implement the 'Grammar for Writing' intervention in 2013;

☐ I agree to providing an information letter to all parents of children in Year 6;

☐ I agree to staff attending professional development days;

☐ I consent to the school taking part in the above study

Name of headteacher
……

Name of School
……

School Tel no
……

Headteacher Email address
……

Name of School Contact (if not headteacher)……

School Contact email address……

Signature of headteacher…..…….Date…..……

Thank you for agreeing to take part in this research. Please return this consent form by post to:

Debra Myhill, Professor of Education, Associate Dean for Research: SSIS, Graduate School of Education, University of Exeter, Heavitree Road, Exeter EX1

Appendix 3: Parent-Pupils Information Letter Year 6

[INSERT DATE]

[INSERT SCHOOL NAME]

Dear Parent / Carer

Your child's school is taking part in Exeter University's Grammar for Writing, which is a new way of teaching writing.

Durham and York Universities are finding out how well the Grammar for Writing programme helps children's writing. This study is funded by the Educational Endowment Foundation, which has been given money from the Government to evaluate new forms of education.

In your child's school, one Year 6 class is using Grammar for Writing and one Year 6 class is having normal teaching. The choice of class is decided by chance, like in a lottery. We will compare results from both classes. We would like to collect some information about your child from your child's primary school. Your child's school will provide information including your child's name, unique pupil number, gender, date of birth, details on your child's current National Curriculum writing level and free school meal status. Your child will do a test designed by GL assessment. When the test is completed the test will be sent to GL assessment for marking, they will send the test results to Durham University and to your child's teacher. We need this information to find out if Grammar for Writing works.

Your child's information will be treated with the strictest confidence. To see if there is a long term impact, named data will be matched with the National Pupil Database and shared between the evaluation team, University of Exeter and Education Endowment Foundation. We will not use your child's name or the name of the school in any report arising from the research. Your child's information will be kept confidential at all times.

If you are happy for your child's information to be used you do not need to do anything. Thank you for your help with this project.

If you would rather your child's school did not share your child's information for this project, please complete the enclosed opt out form and return it to your child's school by [INSERT DATE].

If you would like further information about the Improving Writing Quality evaluation please contact Natasha Mitchell the Evaluation Coordinator: natasha.mitchell@york.ac.uk; 01904 321655

Yours faithfully

Professor David Torgerson (York University)

Professor Carole Torgerson (Durham University)

Professor Debra Myhill (University of Exeter)

Grammar for Writing Evaluation: Opt Out Form

If you DO NOT want your child's data to be shared for use in the Grammar for Writing evaluation, please return this form to your child's school asap.

☐ I DO NOT want my child's data to be shared for use in the Grammar for Writing evaluation

Parent/Carer Signature……
Date……

Child's Name……

Child's School……

Appendix 4: Further Result Details

The Table below contains output from the statistical analyses which was used to calculate effect sizes.

	Coefficient (relating to intervention)	SE of Coefficient	Residual SD (Random Effect)	SE of Residual SD
Cluster analyses				
Primary	0.3369707	0.3270565	3.221293	0.0525334
Secondary (reading)	0.3762526	0.2551135	3.712202	0.0654189
Secondary (spelling)	0.0364394	0.2440797	3.225648	0.0513772
Small group versus control	0.7767005	0.3992479	3.198980	0.0844057
Individual-level analyses				
Primary	0.6741492	0.2812177	3.205891	0.0945098
Secondary (reading)	-0.0285395	0.3659872	3.735150	0.1249519
Secondary (spelling)	0.4841083	0.2984487	3.462565	0.0993129

Appendix 5: Ceiling/Floor Effect Examination

The test and outcomes were assessed for ceiling or floor effects using histograms and summary statistics.

Ceiling effects occur if a test is too easy – i.e. lots of pupils achieve a perfect score.

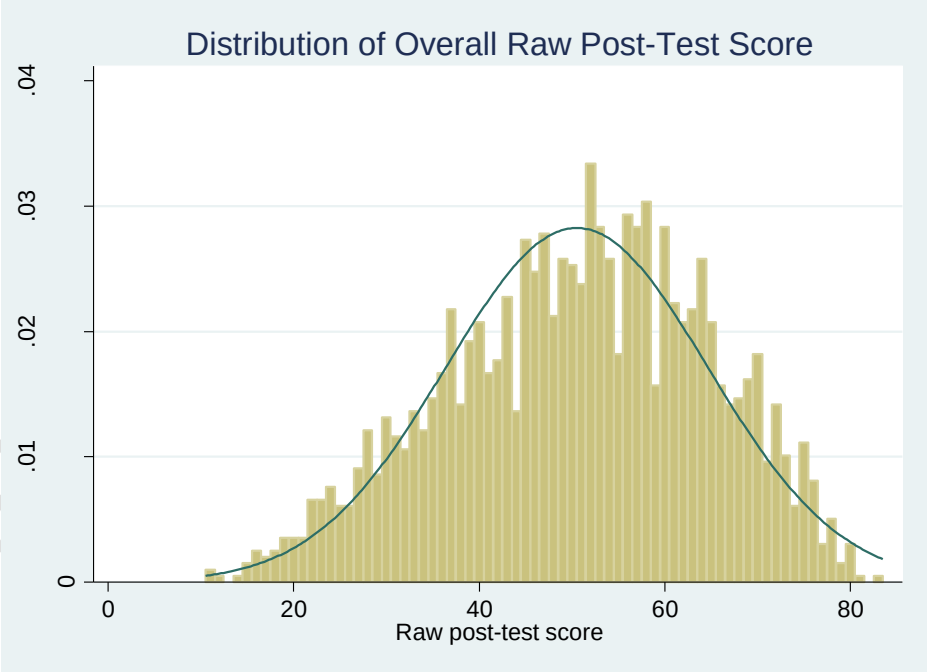
Floor effects occur if a test too hard – i.e. lots of pupils score zero.

There does not appear to be a strong floor or ceiling effect in overall raw score, extended writing, reading or spelling or grammar scores.

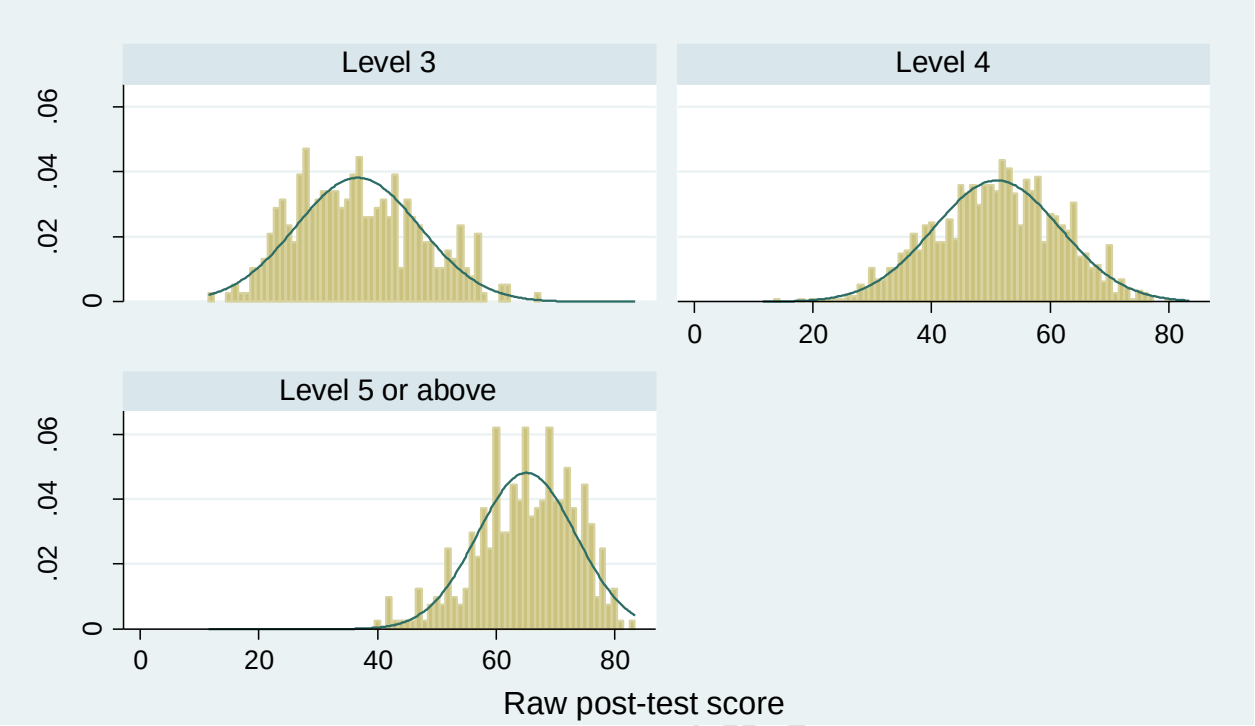
Overall raw score

Overall raw score can range between 0 and 84 marks. The histogram below shows the distribution of post test scores achieved and summary statistics are presented in the Table below. The median and mean scores were similar at around 51 marks which suggests the data are not skewed. The minimum score observed was 11 out of 84: no pupil achieved full marks. The mode score was 52 marks. As seen in the histogram, there was not a large proportion of pupils achieving particularly high or low score implying no ceiling or floor effect.

Overall Raw Score	Min	1st Q	Med	3rd Q	Max	Mean	SD
	11	40	52	61	83	50.5	14.12



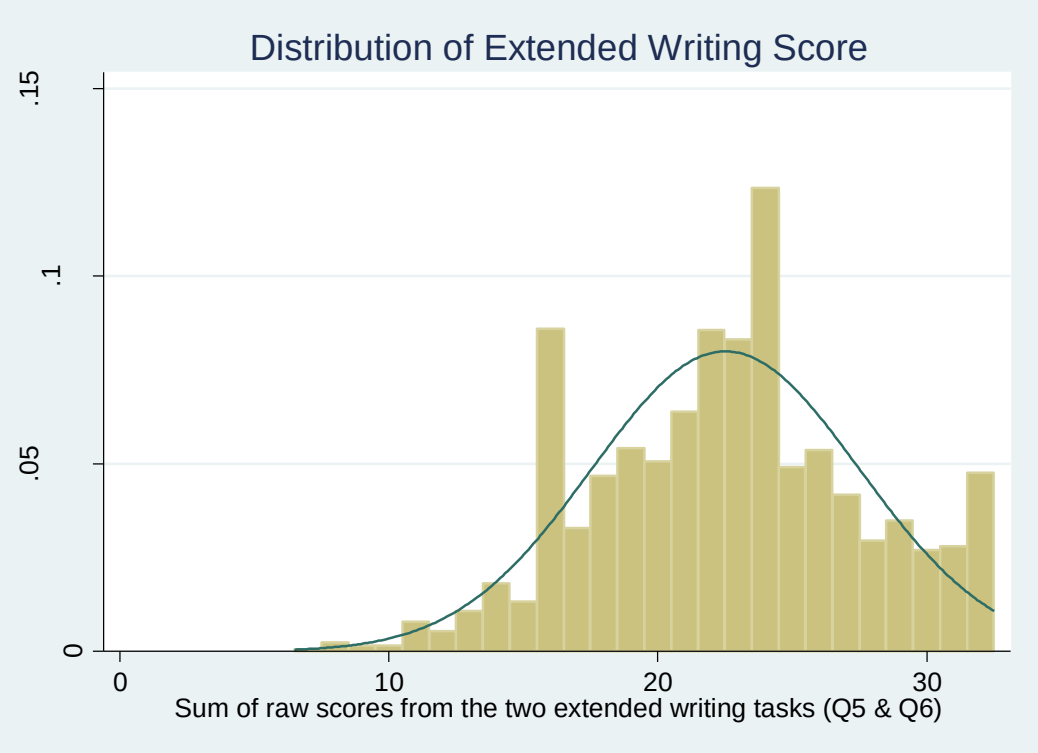
Histograms of overall raw score by predicted level (below) were produced and it appears that the test distinguished between different ability levels.



Primary outcome (extended writing)

The primary outcome of extended writing score can range between 0 and 32 marks. The histogram below shows the distribution of post-test extended writing scores achieved and summary statistics are presented in the Table below. The median and mean scores were similar at around 23 marks suggesting non-skewed data. The minimum score observed was 7 out of 32; the maximum score was 32 marks. The mode score was 24 marks. As seen in the histogram, there was not a large proportion of pupils achieving particularly low score and whilst 97 pupils did achieve full marks, this only accounted for around 5%, implying no ceiling or floor effect.

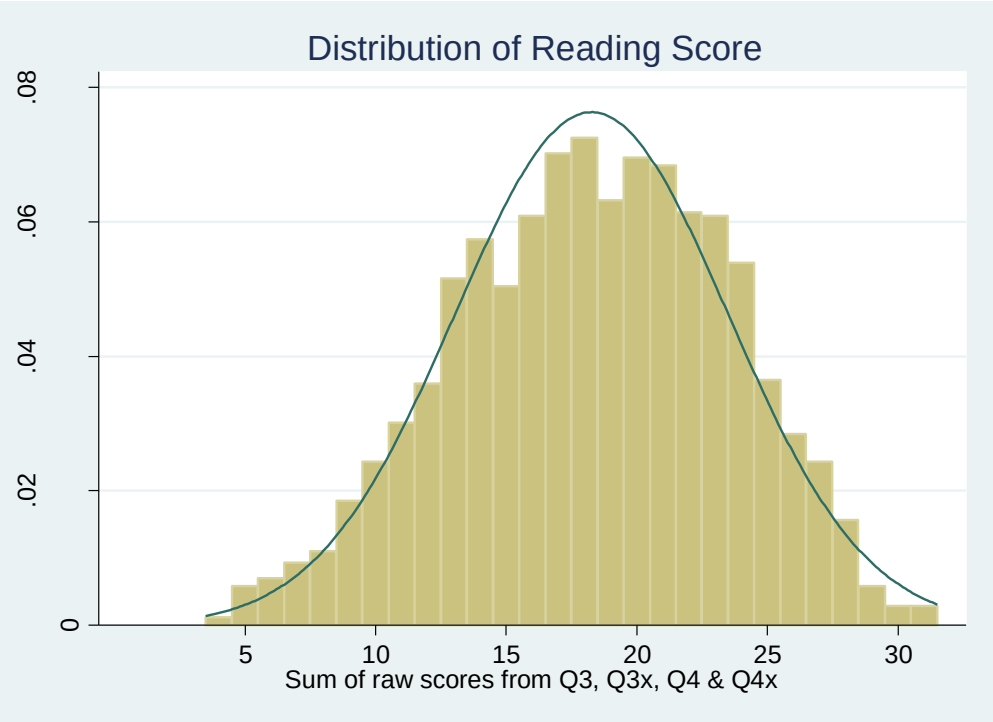
Extended Writing Score	Min	1st Q	Med	3rd Q	Max	Mean	SD
	7	19	23	26	32	22.5	5.00



Secondary outcome – Reading

The secondary outcome of reading score can range between 0 and 32 marks. The histogram below shows the distribution of post-test reading scores achieved and summary statistics are presented in the Table below. The median and mean scores were similar at around 18 marks implying no skew. The minimum score observed was 4 out of 32; the maximum score was 32 [31 in table?] marks. The mode score was 18 marks. As seen in the histogram, there was not a large proportion of pupils achieving particularly high or low score implying no ceiling or floor effect.

Reading Score	Min	1st Q	Med	3rd Q	Max	Mean	SD
	4	14	18	22	31	18.3	5.23



Secondary outcome – Spelling and Grammar

The secondary outcome of spelling and grammar score can range between 0 and 20 marks. The histogram below shows the distribution of post-test spelling and grammar scores achieved and summary statistics are presented in the Table below. The median and mean scores were similar at around 12 marks suggesting the scores do not follow a skewed distribution. The minimum score observed was 0 out of 20; the maximum score was 20 marks. The mode score was 13 marks. As seen in the histogram, there was not a large proportion of pupils achieving particularly high or low score implying no ceiling or floor effect.

[table and text refer to different things]

Extended Writing Score	Min	1st Q	Med	3rd Q	Max	Mean	SD
	7	19	23	26	32	22.5	5.00

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

Education
Endowment
Foundation

The Education Endowment Foundation
9th Floor, Millbank Tower
21–24 Millbank
London
SW1P 4QP
www.educationendowmentfoundation.org.uk