Teaching Grammar and Writing: A Randomised Controlled Trial and Implementation and Process Evaluation of *Englicious*

Dominic Wyse[1], Bas Aarts[2], Jake Anders[3], Alice de Gennaro[3], Julie E. Dockrell[4], Yana Manyukhina[1], Sue Sing[1] & Carole Torgerson[5]

[1] Helen Hamlyn Centre for Pedagogy, UCL Institute of Education, University College London | UK

[2]UCL English, University College London | UK

[3]UCL Centre for Education Policy & Equalising Opportunities, University College London | UK

[4]Department of Psychology and Human Development, UCL Institute of Education, University College London | UK

[5]Department of Education, University of York | UK

Abstract: Very few research studies of the teaching of grammar and writing had been carried out with children younger than eight-years-old prior to the research reported in this paper. The research evaluated a new approach to teaching grammar and writing called *Englicious*. A Randomized Controlled Trial (RCT) and Implementation and Process Evaluation (IPE) research design, featuring 1,246 pupils aged six to seven-years-old in 70 primary school classes, was used to evaluate the effectiveness of *Englicious* for improving children's writing. The approach was implemented in the context of the programs of study for grammar teaching in England's national curriculum. The research found that there was no effect of the grammar teaching intervention on pupils' narrative writing. The effect size for pupils' sentence generation was sufficient to merit reflections about potential impacts of aspects of the intervention although this outcome also did not reach statistical significance. It is hypothesised that the manipulation of words, phrases and sentences, combined with practice at writing, may have contributed to any positive effect, although this would need to be confirmed in future research. It is concluded that until more research is done to investigate the effectiveness of different approaches to teaching grammar and writing with young children, existing evidence-based approaches are more likely to be effective to help young pupils' narrative writing. England's national curriculum specifications for teaching grammar and

journal of
WRITING RESEARCH

Wyse, D., Aarts, B., Anders, J., de Gennaro, A., Dockrell, J.E., Manyukhina, Y., Sing, S., & Torgerson, C. (2025, accepted for publication). Teaching grammar and writing: A randomized controlled trial and implementation and process evaluation of *Englicious*. *Journal of Writing Research*, *volume(issue)*, ##-##. DOI: xx

Contact: Dominic Wyse, Helen Hamlyn Centre for Pedagogy, UCL Institute of Education, University College London, 20 Bedford Way, London, WC1H 0AL | UK - d.wyse@ucl.ac.uk

Copyright: This article is published under Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 Unported license.

writing could usefully be reviewed to more closely reflect the evidence base from this field of writing research.

Keywords*:* teaching grammar, teaching writing, national curriculum

1. Introduction

Helping young children to learn to write is one of the most important goals of elementary/primary education. Pupils in early years settings and schools need to progress from using emergent forms in their earliest attempts at writing towards using the conventional grammar of sentences (Berninger et al., 2011). Requirements for the teaching of writing are specified in the curricula at national, state and other regional levels in many countries of the world.

Debates about the nature of the English language and how it might best be taught are very longstanding, and the teaching of grammar has been an important part of these debates (Wyse, 2001; Wyse, 2017). Research in the last three decades has included studies with findings showing the importance of contextualised grammar teaching in relation to primary/elementary school pupils' word choice errors and the texts they try to write (Wyse, 2006). However, research had not shown a statistically significant positive impact of grammar teaching on students' writing until an experimental trial undertaken with secondary school students showed that an intervention described as contextualised teaching of grammar had a positive impact on secondary students' writing (Myhill et al., 2011). The contextualised teaching of grammar in the intervention drew on systemic functional linguistics. This study was notable because it was one of the first to show this positive impact of grammar teaching on writing. The effective intervention required secondary school students to engage with relatively sophisticated levels of understanding commensurate with the students' ages and stages of development. In more recent work by Myhill (2021) the striking difference between England's national curriculum specification of grammatical terminology for primary students compared with no such specification for secondary students was noted[1].

Understanding some of the differences between how primary students versus secondary students might respond to pedagogies for grammar and writing are an important part of the contribution we make in this paper. It was theorised a-priori that effective teaching of grammar for primary age students may differ from the teaching of secondary students in important ways according to the students' language development, not least for students in the younger age groups (from age five to age seven). Effective interventions for the teaching of reading had shown that more direct forms of contextualisation could be effective with six-year-old children. For example, the contextualised teaching of key language features, such as the relationships between phonemes, letters and words, had included opportunities to put this learning into practice in the same lesson through practising the reading of texts (Wyse & Hacking, 2024). The grammar teaching approach called *Englicious* was built on the theory that

[1] Contrary to the claim of an *a-priori* "binary" position (Myhill, 2021, p. 265), the research published in Wyse (2001) reported an extensive and comprehensive review of empirical studies available at the time. Peer-reviewed confirmation of the extensive nature of the paper can be seen in the systematic review by Andrews et al. (2004). The research emerged from questions raised by a multidisciplinary team of researchers, teachers, and a local education authority lead.

an intervention that blended the teaching of key grammar features in a relatively decontextualised way followed by the opportunity to put this into practice through contextualisation in narrative writing in the same lesson might be effective, and was an approach that had not previously been subject to robust evaluation.

The *Englicious* teaching approach was designed by the linguists on the research team. It was built in part on their research on corpora of the English language (Aarts, 2011; Aarts et al., 2016; Aarts & Smith-Dennis, 2018) but also through the development of a website used at the time of the research by more than 10,000 teachers (*Englicious* Survey of English Usage, 2024). The *Englicious* intervention was informed by the hypothesis that understanding of the differences between grammatical form and grammatical function may be an important element to help children learn to form grammatical sentences in writing (Aarts, 2019). To some degree the emphasis on form and function in the study links conceptually with other theories of language in use, and language as a system (e.g. see Fontich et al., 2020). It was hypothesised that explicit teaching of the grammar concepts required by England's national curriculum could be systematically linked to pupils' writing, within the same lesson, to enhance pupils' writing. The key elements of the study's manualised teaching intervention included a blend of decontextualised and contextualised teaching; explicit explanations about grammatical terms; interactive experimentation with combinations of words and sentences; and overall, a descriptive rather than prescriptive approach to grammar teaching (see Appendix for example from the teaching manual used in the intervention).

2. Studies of grammar teaching

Research evidence on effective teaching of grammar in relation to teaching writing at elementary/primary education level has been summarised in systematic reviews (SRs), meta-analyses (MAs), and tertiary reviews (TRs). An early SR by Andrews et al. (2004) that included studies focused on the teaching of primary age pupils, carried out by a team based in England, addressed syntax defined as the "internal elements of the sentence, classifying 'parts of speech' and describing (and sometimes prescribing) the relationship between parts of speech" (Andrews et al., 2004, p. iv). The reviewed research studies featured a range of grammar interventions including *generative grammar*, *transformational grammar* and *traditional grammar*. The review concluded that the teaching of grammar had virtually no impact on pupils' writing, and, as a result, they argued it should not be part of the national curriculum in England.

Subsequently a MA addressed "the explicit and systematic teaching of grammar (e.g., the study of parts of speech and sentences)" (Graham & Perin, 2007a, p. 449). Overall, the results indicated grammar teaching had a statistically significant negative effect on pupils' writing. In a TR published in the same year as the MA, Graham and Perrin (2007b) added analysis of single subject design studies, and recurring themes from qualitative studies. The conclusions of this review supported previous findings about the negative impact of grammar teaching for writing but also found that there was tentative evidence for the impact of *sentence combining*. Sentence-combining involves teaching students to construct more complex and sophisticated

sentences through exercises in which, for example, two or more basic sentences are combined into a single sentence. Positive effects for sentence combining have also been found for struggling writers in primary schools (Walter et al., 2021).

A more recent SR and MA was that carried out by Graham et al. (2023) addressing students in grades six to 12 (aged 12 to 18). Contrary to previous SRs and MAs of the lack of impact of grammar teaching on secondary pupils' writing, Graham et al. (2023) found a positive effect size of 0.77 for all writing outcomes across the selected studies of grammar teaching (the inclusion of the Myhill et al. (2011) study had an impact on this effect size). The disparity with the negative findings of Graham & Perin (2007a) about grammar teaching was hypothesised to be that for the purpose of analysis of effects the previous MA had treated control conditions that included grammar teaching as experimental conditions.

Although the SRs cited so far in this paper included some attention to primary age pupils, the first MA on teaching writing to focus exclusively on studies carried out with elementary/primary school students was carried out by Graham et al. (2012). It was found that teaching grammar did not have a statistically significant positive effect on writing quality; indeed, half of the study effects were negative. It was concluded that there were other evidence-based approaches to teaching writing that were more likely to be effective - confirmed also by Koster et al. (2015). The authors' caveats about their findings included noting that only four studies of grammar had been undertaken with elementary pupils, and that overall, the quality of the studies was not optimal. For example, only 36% of all the studies included in the MA had used random assignment of pupils to conditions: non-random assignment could have introduced selection biases. Attrition and ceiling and floor effects problems were also common in the studies, something which increases the likelihood of measurement inaccuracies. And fidelity to the intervention had rarely been established in the studies, as such the measured effects may not have been due to the intervention as intended.

An RCT of grammar teaching with primary school students, adopting the contextualised teaching of grammar approach of Myhill et al. (2011) that included intention to treat analysis and attention to fidelity of interventions, was completed after the SR and MA by Graham et al. (2012). The cluster RCT by Torgerson et al. (2014) covered pupils in England aged 10 to 11 and aged 11 to 12. The impact of contextualised grammar teaching was small and did not reach statistical significance (for further reflections see Wyse & Torgerson, 2017). Although Torgerson et al. (2014) included primary age pupils these were still not the younger age primary students that our paper is focused on.

This paper's review of SRs and MAs, and some individual studies, shows the lack of research studies providing evidence of what might be effective for primary age pupils in relation to teaching grammar for writing. In addition to methodological quality issues in some of the studies included in the MAs above, a further limitation in the field has been the absence of studies investigating the impact of writing interventions during the early stages of pupils learning to write, particularly pupils younger than age eight. Young pupils' ability in sentence-combining is one predictor of writing proficiency: the ability to combine sentences that use conventional grammar develops at around the age of seven for most pupils (Berninger et al.,

2011). Consideration of the initial stages in the development of written text production offers the opportunity to examine the ways in which pupils' competence in the grammar of writing can be developed.

In general, the number of studies of grammar and writing with pupils at elementary level is very low. Apart from the approach to teaching grammar known as sentence-combining there were no RCT studies that had shown improvements in pupils' writing as a result of using an explicit approach to teaching grammar. The study reported in this paper was the first to adopt a combined RCT and Implementation and Process Evaluation (IPE) design on grammar teaching and writing with children aged six to seven.

The research was also the first study with a RCT design to evaluate a key component of England's national curriculum programs of study for grammar and writing that were implemented from 2014 onwards (Department for Education (DfE), 2013). England's national curriculum of 2014 included for the first time much more attention to the explicit teaching of grammar, including grammatical terms, in its programs of study. For example, the national curriculum included a five-page appendix of vocabulary, grammar and punctuation content that had to be covered including, for children aged six to seven, introduction to ideas such as progressive forms of verbs, apostrophes, and the grammatical terminology of noun phrase, tense, compound, command, etc.

In order to address the gap in the field of studies about teaching grammar and writing for young pupils the research questions that informed the study reported in this paper were as follows:

1. To what extent is the grammar intervention *Englicious* effective in improving pupils' writing?

2. What are the main implications for teacher practice as a result of implementing *Englicious*, and, more generally, for evidence-informed teaching of writing?

The research sub-questions for the IPE part of the research were as follows:

1. What are teachers' beliefs, knowledge and experience of teaching grammar for writing?

2. To what extent does the teaching in the intervention classes show fidelity to *Englicious*?

3. In what ways does the teaching in the intervention classes differ from the control classes?

4. How do Year 2 teachers deliver the requirements of grammar in England's national curriculum?

3. Methodology

The hypothesis of the research was that the *Englicious* intervention that involved explicit grammar teaching combined with practising writing would improve six-year-old and seven-year-old pupils' sentence generation and narrative writing more than the varied teaching of grammar in 'business as usual' control classes. The research design involved a RCT and an IPE. The RCT was an experimental manipulation with participants randomised to either the grammar intervention called *Englicious*, or to a business-as-usual control in which the teachers followed their standard procedures for teaching grammar and writing.

3.1 Sampling

The target population for the study was all Year 2 teachers and pupils in their classes (aged six to seven) in all state-funded primary schools in a large city in the South of England, excluding special schools that serve pupils who require specialist support services. Schools were initially selected at random from the UK Government's Department for Education list of all schools. Final selection of schools for the sample required additional communications with contacts, including in two counties adjacent to the city that was the site of the study, to ensure sufficient numbers of teachers were included in the study in relation to the power calculations (see below). An evaluation protocol and subsequent statistical analysis plan were published in protocol documents prior to completion of the research (Anders et al., 2021; Anders et al., 2019).

Preliminary power calculations for what was a stratified school-level cluster trial randomisation indicated that for the target sample of 60 schools the following assumptions would be sufficient to detect a minimum effect size of 0.24:

- 15 pupils tested per school (giving a generous allowance for data processing objections and non-response, as part of power calculations);
- intra-cluster correlation of the outcome measure of 0.15;
- 0.49 of post-test variance (corresponding with test-retest correlation of 0.7) in outcome explained by model covariates at both individual- and cluster-level using four regressors;
- two-tailed significance tests at 0.05-level and power of 0.8.

Seventy schools and Year 2 teachers agreed to participate in the trial, met the eligibility criteria, and provided the necessary data for randomization to proceed (further information below). The mean percentage of pupils qualifying for Free School Meals (FSM – a measure of deprivation in the UK) was close to the national average (see Table 1, and https://explore-education-statistics.service.gov.uk/find-statistics/school-pupils-and-their-characteristics).
The proportion of pupils with English as an additional language was approximately double the national average (see Table 1 and https://explore-education-statistics.service.gov.uk/find-statistics/school-pupils-and-their-characteristics), reflecting the geographical focus of the sample.

The flow of participants through the study is detailed in Figure 1.

3.2 Design and random assignment method

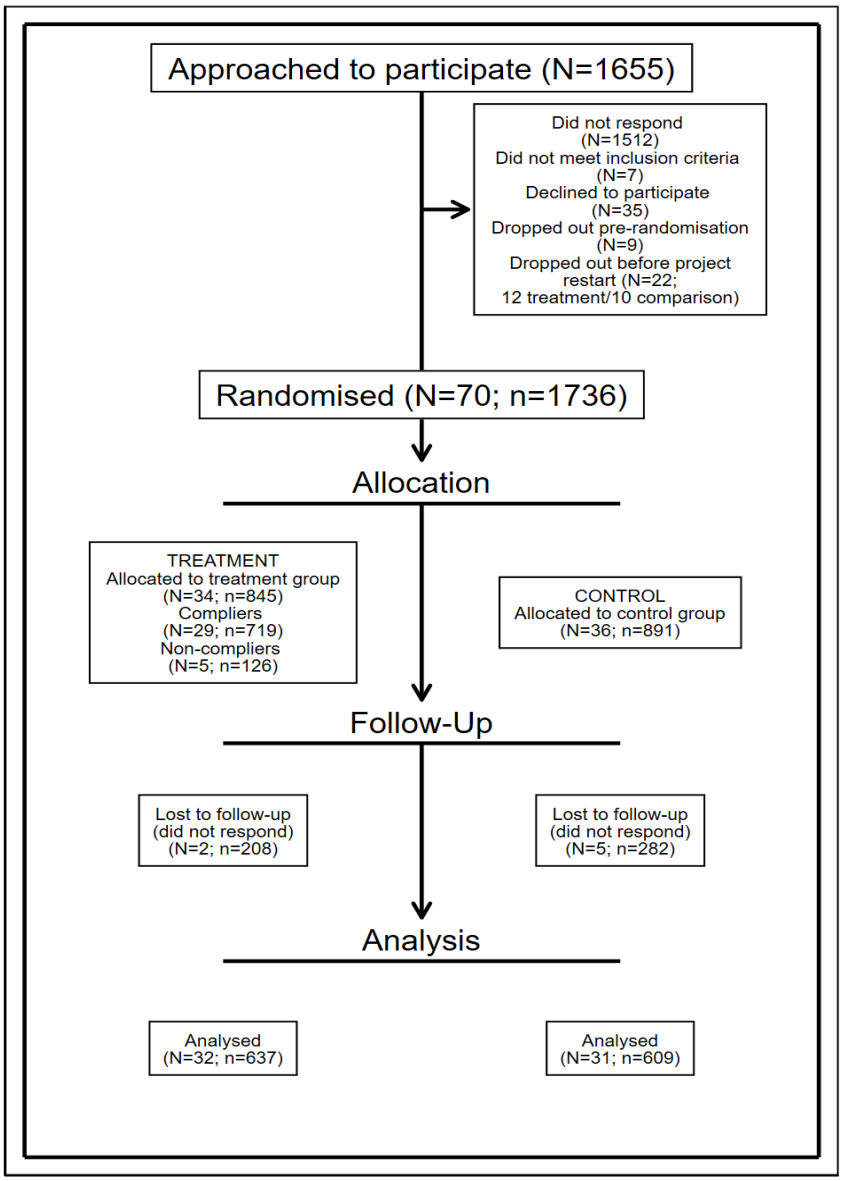
Randomisation was carried out at the school/teacher-level (only one teacher was recruited within each school, hence these two are indistinguishable for research design purposes). Randomisation was always planned to be carried out in two batches for reasons of delivery and recruitment practicalities, with the aim being that each batch be the same size. 60 Year 2 teachers were to be randomly allocated to two groups: *Englicious* Intervention and waitlist control (who received professional development in the intervention after data analysis had been concluded), in equal proportions, resulting in the following batches:

- Randomisation Batch 1 (November 2019): 24 schools were allocated in equal proportions (12 to treatment; 12 to control)
- Randomisation Batch 2 (January 2020): 40 schools were allocated in equal proportions (20 to treatment; 20 to control)

This provided an initial sample of 64 schools, slightly above our 60-school recruitment target. We moved ahead with intervention delivery in school term Spring/Summer 2020 based on these allocations. However, COVID-19 disruption meant that the intervention was abandoned, and delivery re-started in Summer term 2021. This meant that the planned sample of pupils would all be too old to take part in the planned delivery and, in a substantial number of schools, the participating Year 2 teacher would not necessarily be teaching Year 2 in Summer 2021. This change, and the COVID-19 disruption more generally, meant that 22 schools across Batch 1 (12 treatment; 10 control) decided not to take further part in the project.

Re-randomisation of schools who remained on the project after the Covid-19 pause was rejected on the grounds that teachers allocated to the treatment group had received training and so it would not be realistic to treat them as true comparators if allocated to the control group in a randomisation. As such, remaining Batch 1 and 2 schools were retained in their allocated treatment group. Further recruitment was carried out in in Autumn 2020 and baseline testing (see measures section below) in Spring 2021 resulting in the following groups and allocations:

- Batch 1 (updated): 18 schools in equal proportions (9 treatment; 9 control)
- Batch 2 (updated): 24 schools with slightly fewer treated schools remaining (11 treatment; 13 control)
- Batch 3 (March 2021): 28 schools were allocated in equal proportion (14 to treatment; 14 to control)
- **Overall**: 70 schools (34 treatment; 36 control)



Adapted from: www.consort-statement.org/consort-statement/flow-diagram

Figure 1. Participant flow (CONSORT) diagram.

The schools/teachers in the sample were randomly allocated either to the intervention or the control group. At the point of the trial restart post-COVID-19 randomisation, 845 pupils in 34 schools were allocated to the intervention group, and 891 pupils in 36 schools to the control group.

Randomisation within each batch was, as reported in the pre-published protocol, carried out within stratification blocks to reduce the risk of imbalance on important characteristics between our resulting treatment and control groups. These stratification blocks were formed by the intersection between equally sized high and low EAL proportion, and high and low FSM proportion groups.

3.3 Procedure and materials

In addition to the potential importance of an explicit focus on grammar teaching combined with practising writing in the same lessons (see rationale in the introduction to this paper) a further rationale for the intervention was that explicit teaching of grammar was a statutory part of the programmes of study for writing composition in England's national curriculum and therefore had to be delivered by primary school teachers. The national curriculum required children in Year 2 to be taught an extensive set of grammar concepts and terminology (Department for Education, 2013) including all the terms addressed below in the 10 lessons of the intervention.

The *Englicious* intervention was manualised for this research by four teachers of Year 2 children not otherwise involved in the research study, with the support of the linguists on the team (see Appendix for an example lesson. The corresponding author can provide a copy of the intervention manual, which includes the instructional materials).

Teachers in the intervention group received one full day of in-person professional development explaining the rationale for the *Englicious* intervention and how it was to be implemented. This training was led by the linguists on the research team who had developed *Englicious*. A PDF manual for the intervention was explained and given to the teachers in hard copy, and access to digital copy. Teachers also were given access to the relevant *Englicious* resources such as PowerPoint slides and a link to the *Englicious* website. Contact details of a researcher on the research team were given in case teachers had further queries.

The main focuses of the 10 one-hour weekly lessons in the intervention were as follows:

- Lesson 1: Nouns. To understand that nouns are words that represent people, places and things, to understand how they are formed, and to understand what some of their grammatical characteristics are.
- Lesson 2: Adjectives and expanded noun phrases. To understand the grammatical characteristics of adjectives and expanded noun phrases.
- Lesson 3: Verbs. To understand what verbs are, how they are formed and their grammatical characteristics.

- Lesson 4: Adverbs. To understand what adverbs are, how they are formed and their grammatical characteristics.
- Lesson 5: Present tense. To understand, identify and apply the present tense, including the present progressive form of the verb.
- Lesson 6: Past tense. To understand, identify and apply the past tense, including the past progressive form of the verb.
- Lesson 7: Sentence patterns. To understand the grammatical characteristics of the sentence patterns 'statement, question, command, and exclamation' in English, and how they are used.
- Lesson 8: Linking (1). To understand how we can link words, phrases and clauses using coordination.
- Lesson 9: Linking (2). To understand how we can link clauses using subordination.
- Lesson 10: Consolidation. To consolidate and revise existing grammatical knowledge.

All the intervention lessons had a similar structure that included materials on Interactive Whiteboard (IWB) slides. The first three elements of the lesson involved whole class teacher-pupil interaction led by the teacher. The final element of the lessons required the pupils to carry out independent writing although normal peer-to-peer interaction during the activities was permitted. The four elements of the lessons were as follows: 1. 'starter' activity designed to initiate discussion of a particular topic (10 minutes), e.g., in a discussion of nouns, pupils were presented with labelled images of objects and asked, 'What kinds of things do these words represent?'. 2. The teacher would then explain a particular grammatical term using age-appropriate examples, illustrated with images (10 minutes). 3. The pupils' learning was consolidated using practice material from *Englicious* (10 minutes). 4. In the final part of each lesson pupils were asked to carry out a short piece of writing using the grammar they had been taught, which was then shared and discussed in the class as a whole (20 minutes). Lessons ended with a short summary of what had been learnt (10 minutes).

Teachers in the business as usual control condition were asked to continue their usual practices for grammar teaching. As detailed above, teachers of Year 2 children (aged six to seven) in primary schools in England must cover the specifications for grammar knowledge in England's national curriculum, but they have a relatively high level of autonomy to decide their approach to teaching grammar. Considerable variation in pedagogy of teachers in the control group was expected. This variation includes the extent to which any grammar teaching would be contextualised, and the amount of time devoted to teaching writing more generally. The main elements of the grammar and writing teaching approaches that our sample of control group case study teachers carried out is presented at the end of the IPE results section of this paper.

3.4 Measures

Exploration of some of the links between the teaching of grammar and the development of young children's writing was at the heart of the study, and this included the selection of measures to evaluate the effectiveness of a new grammar intervention through tests of writing that assessed narrative writing and the ability to generate sentences, both measures that had been correlated with the development of writing (Abbott and Berninger, 1993; Drijbooms et al., 2017; Kim and Schatschneider, 2017). The ability of pupils to generate more text, e.g., more grammatically conventional sentences, is one component of pupils' development of writing (Kim & Schatschneider, 2017) and can be considered a foundational skill for producing extended texts such as narratives (Dockrell et al., 2019). Two measures of grammar and writing were chosen, consistent with the hypothesis of the research that grammar teaching could improve pupils' writing. The measures were also chosen because they were appropriate to the age of the pupils involved, and were likely to avoid ceiling or floor effects to ensure that there were items that all pupils should be able to make at least a good attempt at.

1. Progress in English (PiE) test: the primary outcome measure. A standardised measure to assess pupils' writing through a narrative writing test. We secured agreement from *GL Assessment* to use the element from their Progress in English (PiE) test that was focussed on writing. GL Assessment's more recent Progress Test in English (PTE) did not include a standardised writing element which is why the PiE was used. The baseline task, used as the pre-test, was the PiE short task which required the pupils to write some instructions for growing plants. Some picture cues were provided in the answer booklet. The pupils were allowed five minutes to plan their writing plus 20 minutes to carry out the writing. The long version of the PiE was chosen as the primary measure and post-test for the research because the key research question was whether grammar teaching could improve pupils' writing. The long version PiE test required pupils to produce a piece of narrative writing based on a brief discussion with their teacher about the prompt which was to write about a rainy day out with their family. The teacher introduced the writing task to the whole class. The pupils were allowed a maximum of 10 minutes thinking/planning time plus 40 minutes for writing which was done independently and written onto a preprepared writing booklet. There were no text length requirements.

2. Sentence Generation Test (SG): the secondary outcome measure. A bespoke measure to assess grammar through pupils generating sentences from word prompts. The SGT was based on a test derived from a previously published study (Arfé et al., 2016). The SGT was chosen to detect an impact on pupils' sentence-level skills, considered a precursor to text level production in younger pupils. This measure of pupils' writing competence relates to fluency in writing sentences, and the grammatical and semantic appropriateness of the sentences that pupils generate. The teacher offered two words that the children had to combine in order to write as many different sentences as possible for a maximum of five minutes. This was followed by a second pair of words. The same word pairs were used at baseline and immediate post intervention.

Pupils were tested twice on the writing outcome measures: 1. Baseline: prior to the start of the intervention; 2. Post-intervention: one to two weeks after the end of the intervention. As described above the baseline task for the PiE was the short task; the post-intervention was the long task. The baseline and post-intervention test for the SGT was the same task. We planned for additional delayed tests but due to COVID-19 this was not possible, hence a deviation from the published protocol. Due to alterations to delivery of the trial caused by COVID-19 restrictions, both outcome and pre-test measures were administered by classroom teachers, through whole class implementation, as opposed to researchers. This meant that administration of tests was not blinded to treatment allocation. However, the nature of the writing tests, and the scripting of the introduction to the tests given by the teachers, minimised the potential effects of this on test performance among pupils.

3.5 Test Marking

The marking criteria for the writing task from PiE were: writing composition and effect; text organisation; sentence structure; and vocabulary. The individual criteria were scored separately (composition and effect 0 to 8 marks; text organisation 0 to 4; sentence structure 0 to 4; vocabulary 0 to 4) then added for an overall score per pupil from zero to 20 marks (the approach to marking is covered below). This test's marking criteria had been developed with relevance to England's national curriculum. For the PiE measure the reliability was found to be greater than 0.9 (Cronbach's Alpha; Kirkup et al., 2006).

For the SG test one mark was awarded for each unique sentence a child produced for the test (zero marks in total if the sentence was not different from previous sentences); one mark was awarded if the sentence was written using standard English grammar; and one mark was awarded for semantic meaning (if the sentence made sense on its own). Inter-rater reliability for the sentence-generation task had previously been found to be good (greater than 0.95; Walter et al., 2021) and test-retest reliability at a two-month interval was 0.62 (op. cit.).

All marking for the project was carried out by 23 markers with sufficient expertise in the teaching and assessing of writing who were specially recruited, via open application and interview, to the project. Markers received training, given by the full-time researcher and by the Principal Investigator (as senior raters) on the marking of all tasks to ensure consistency in their approach. The training was a one-day live online session. The approach was set out in a Marking Guide which was shared with all markers. This was followed by practice marking of 4-5 scripts which were not part of the trial sample. These scripts were analysed for agreement and error patterns, including calculation of an inter-rater reliability statistic compared to the agreed mark by senior raters. Our agreed procedure was that if the inter-rater reliability of a marker's scripts at this training stage was below 0.6 then further training would be carried out with a further 4-5 scripts analysed in the same way, however no markers fell below this standard at the training stage.

For the main marking task of the research, markers were kept blinded from whether any given test they were allocated was treatment or control, since this would have had a potential to introduce bias. Markers were allocated a mix of tests from treatment and control groups to

reduce risk that tester effects could drive results at the margin. Markers marked in batches of different tasks to prevent the possibility of their perception of one task shaping their marking of another task, particularly across pre- and post-tests.

For the main marking task a 20% sample of the first 100 scripts of each task marked by each marker was also marked by a second marker from the team of markers. If inter-rater reliability of a marker's sampled scripts was above 0.7 then the checks moved to double marking 1% of their remaining scripts for monitoring. However, if the inter-rater reliability of their scripts fell below 0.7 then a senior rater investigated, and further training was provided for the markers as appropriate. This happened for seven markers, with the additional training following the approach above. The retrained marker then repeated the initial 20% sampling process. All but one of the markers identified by this process had inter-rater reliability above 0.7 in these supplementary checks, meaning we then also moved to a 1% sampling of further scripts for double marking to monitor the remainder of the marking process. In the case of the one marker of the seven retrained markers the quality of marking was such that it was decided to end their involvement in the marking exercise with all of their scripts reallocated to other members of the team and fully re-marked. Analysis of the full set of scripts that were marked at least twice provided average inter-rater reliability (IRR) estimates for the PiE of 0.72 and for the SGT of 0.85. These should provide lower bounds of the reliability of the marking since the double-marked sample is skewed towards markers identified as having discrepancies with others.

3.6 Implementation and Process Evaluation Data Collection

The IPE design consisted of three surveys of the teachers in the study, and 12 case studies of the work of teachers. The surveys were: 1. pre-intervention survey – sent to all teachers; 2. mid-intervention survey – sent only to intervention group teachers; 3. end of intervention survey - intervention group teachers. The mid-intervention and end of intervention surveys were primarily aimed at acquiring data about teachers' perceptions about teaching England's national curriculum grammar and writing requirements, and evaluation of the implementation of the *Englicious* intervention including fidelity to it. The pre-intervention survey included questions about the teachers' experiences of teaching; their professional development in grammar teaching and writing teaching; other sources of information about teaching; their perceptions of their effectiveness as teachers of grammar and writing; and the frequency they used some key elements in their teaching of writing.

For the case studies, 20% of the teachers in the intervention and control groups were selected at random: six intervention schools and six control schools. Two visits were made to each Year 2 teacher's class to observe a grammar lesson and to interview the teacher: once at the beginning of the 10-week intervention period and another towards the end of the intervention period. In the pre-COVID-19 part of the research, the school visits were completed by four members of the research team working in pairs in the first visits to all schools in order to combine different kinds of expertise and to agree interpretations of what was seen. Subsequent visits were carried out by individual research team members.

The first visit consisted of the following elements: 1. an overview interview to collect relevant baseline information about the teacher and the school (approximately 20 minutes), and in the case of control schools to elicit their approach to teaching grammar; 2. an observation of at least one full lesson where the intervention, or other grammar teaching in the control schools, was the focus; 3. a concluding interview (approximately 30 minutes). This interview included the use of examples from the observed teaching as a stimulus to elicit greater depth of understanding about the implementation of the intervention or the grammar teaching in the control schools. In the intervention schools, the interview also elicited early reflections about the intervention.

The second visit consisted of 1. an observation of at least one full lesson where the intervention or other grammar teaching was the focus; 2. a concluding interview using examples from the observed teaching as a stimulus to elicit greater depth of understanding about the implementation of the intervention or the grammar teaching in the control schools (approximately 30 minutes).

The first visit and second visit interviews were structured according to questions on interview schedules. These schedules were designed to ensure as much information about the intervention teaching and the control teaching was elicited. Topics for interview questions included the teachers' approach to teaching grammar; knowledge about grammar; professional development received; planning done to inform grammar teaching; and teachers' views on England's national curriculum requirements. The interviews were audio-recorded and fully transcribed.

Observations of lessons were first recorded as field notes (handwritten and/or using digital devices), and then transferred to an observation proforma as soon after the observation as possible. The proforma itemised the following areas and therefore served to direct the observational focuses: implementation environment, implementer characteristics, and participant characteristics. The areas were also reflected in questions that acted as prompts for the observations, e.g. to what extent does the teacher's delivery reflect the intervention? What is grammar teaching like in the control group? What is the nature of the classroom environment and ethos including physical characteristics? In what ways does the teaching reflect the objectives of the lesson, as part of the intervention more broadly?; What is the nature of the pupils' response to the teaching?

Validity and reliability of the classroom observations was enhanced through the qualitative methodology of the IPE of the research project. All observers of lessons had strong understanding of primary literacy teaching for example as a result of prior experience as being teachers. The early observations that involved two team members were discussed by the pairs of observers. Interpretation of the lessons that they both observed were agreed, and the information added to lesson observation proformas. Subsequent case study data from each visit by the full-time researcher was reviewed and interpretation agreed through discussions by the full-time researcher and the PI of the project. All findings were also agreed by the whole research team as a result of the writing of the main report and other publications, and through review by an advisory group who the research team met with three times.

3.7 Fidelity of implementation

At the one-day *Englicious* intervention training, teachers were told that the lessons had been designed to include a small margin of flexibility to account for teachers' professional judgements, e.g., identifying when pupils might need additional explanations for grammatical terminology, but in the main they should do their best to stay as close as possible to the lesson plans. After each grammar lesson the teacher made brief notes in the teaching manual, including the duration of the grammar lesson, which was regarded as one measure of fidelity. Lessons that had fidelity to the intervention were deemed to be between 45 minutes and 75 minutes.

We present some final reflections about fidelity in the results section of the paper, but we present some initial information here. Of the 17 mid-intervention survey responses to the question about implementation fidelity, ten respondents reported that implementation was "fully as intended, to the best of my knowledge", six respondents selected "mainly as intended", and just one selected "quite a lot of changes from intended"; the remaining two respondents to this survey did not provide an answer to this question.

As part of the data collection for the IPE case studies of teachers the research team observed 11 intervention lessons. The shortest observed lesson lasted 54 minutes and the longest lesson was 75 minutes. All 11 lessons that were observed as part of the IPE school visits were deemed overall to show fidelity to *Englicious* in terms of lesson duration and to the approach and lessons in the intervention manual. The analysis of IPE data for fidelity found four more notable deviations from the manual all of which were deviations in duration of the parts of the structure of lessons (see below in results section).

In terms of fidelity, we also used lesson duration to explore whether compliance with the intervention as intended, proxied by measuring delivery of all lessons of the intended duration, affected the estimated impact of the intervention using a complier analysis discussed in the results section. For the purposes of that complier analysis, we calculate two statistics from these measures specifically:

1. As a continuous measure/implementation index, the number of classes in the sequence of the appropriate length (standardised to have mean zero and standard deviation one to aid interpretation);
2. As a binary measure, schools in which teachers report having successfully delivered all classes in the sequence of the appropriate length will be deemed to be compliant and those who do not report this are deemed non-compliant. 22 out of 32 treated schools (representing 447 out of 637 pupils in the treatment group) were deemed to be compliant by this measure.

4. Data analyses

The procedures for the quantitative data analyses of the test measures were pre-specified in a statistical analysis plan (SAP) (Anders et al., 2021). Imbalance in baseline measures and characteristics is reported using standardised differences (Imbens & Rubin, 2015) — often referred to as 'effect sizes' but without a causal interpretation in this context.

We primarily estimate the impact of the intervention on an intention to treat (ITT) basis, i.e., including all of those assigned to the treatment group in our analysis regardless of whether or not they fully implemented the intervention as intended (as described above). This is typically held to be a more reliable guide to the likely impact of implementing an intervention, given that compliance with an intervention is required as part of agreeing to be part of a study and, if anything, likely to be worse outside the context of a research study. We will go on to estimate the impact among those who fully complied with the intervention below.

The average intention to treat effect of the treatment on our primary and secondary outcome measures are estimated using hierarchical (multilevel) linear regression models of the form:

$$Y_{ij}=\alpha+\beta_1 Treat_j+\gamma X_{ij}+\eta_j+\lambda_j+\varepsilon_{ij}$$

where Y_{ij} is the outcome measure (long PIE score in the primary measure analysis; sentence generation task score in the secondary measure analysis) for individual i in school j, $Treat$ is our school-level treatment indicator, X is a vector of school- and pupil-level covariates (including our baseline test measures, gender, eligibility for free school meals as a proxy for low income, an indicator of having English as an additional language, as well as group-level averages of these measures) to improve precision, η_j is a vector of randomisation stratification variables to align the model with the randomisation design, $\lambda_j \sim N(0, \sigma_\lambda^2)$ is a school-level random effect assumed to be normally distributed and uncorrelated with the pupil-level error term ε_{ij}. The primary estimate of interest is the β_1 coefficient on the $Treat$ indicator and it is this that we focus on in reporting, also converting the raw estimate into a Cohen's d effect size using the standard formula (e.g. Fritz et al., 2012, p. 7).[2]

Given the nature of the data, it is important for our statistical inference to take the clustering of pupils into schools into account. We do so using our multilevel modelling approach.

Estimates for pre-specified sub-groups were estimated using an interaction of the treatment and an indicator variable for the sub-group of interest:
$Y_{ij}=\alpha+\beta_1 Treat_j+\beta_2 SubGroup_{ij}+\beta_3 Treat_j * SubGroup_{ij}+\gamma X_{ij}+\eta_j+\lambda_j+\varepsilon_{ij}$ where $SubGroup_{ij}$ is an indicator variable for the sub-group of interest. In other respects the specification and method of analysis remains the same as the main analysis model reported above. The primary estimate of interest is β_3 (i.e., on the interaction between the treatment and the sub-group of interest), which recovers the difference between the overall treatment effect and the sub-group treatment effect. We also report $\beta_1+\beta_3$ which recovers the treatment effect within the sub-group, which we additionally report after conversion to a Cohen's d effect size.

To understand whether issues of imperfect compliance to the intervention may have affected the primary intention to treat estimates, we conducted complier analyses in two

[2] Note that this modelling strategy differs from our pre-specified analysis strategy (Anders et al., 2021) based on reviewer and editorial guidance. However, it is highly similar to the estimates from our pre-specified analyses, as reported by Wyse et al. (2022).

ways. First, using a complier average causal effect (CACE) analysis (Gerber & Green, 2012) estimated using a two stage least squares (2SLS) regression approach for a binary indicator of full compliance (this will produce an estimate of the treatment effect for those who fully complied with the intervention). Second, using an instrumental variables analysis, again estimated using a 2SLS regression approach, with a continuous measure of implementation scaled such that we obtain a treatment estimate associated with a one standard deviation increase in the number of lessons in the sequence delivered.

Further details about the methods and code used for all of the following analyses are available online (Anders et al., 2021): balance of sample analyses; primary outcome (PiE test) analysis; secondary outcome (sentence generation) analysis; sensitivity analysis, of treatment indicator and covariates; robustness analysis to explore potential effects of pause and restart due to COVID-19; graphical analyses; sub-group analyses; missing data analysis; fidelity/compliance analysis (more than plus or minus 15 minutes intervention lesson time); and effect size calculations.

IPE quantitative survey data were analysed using descriptive statistics. Qualitative survey data, interview transcripts, and lesson observation fieldnotes, that had been transferred to lesson observation proformas, were subject to qualitative data coding and the development of overarching themes. A qualitative data analysis (QDA) plan was developed. The QDA analyses were led by the full-time researcher on the project supported by the PI. A combination of deductive and inductive coding was used consistent with the interpretivist orientation to qualitative data analysis (Atkinson & Delamont, 2005; Kawulich, 2017). The QDA of observations and interviews data began with a reading of all data files in full (transcriptions of interviews; lesson observation proformas) and annotating with potential qualitative codes, informed by the main topics of the IPE research questions. Validity of QDA was enhanced through meetings between the full-time researcher and the PI of the project to agree interpretations of data coding. These were subsequently reviewed by the wider research team at various points, particularly during the drafting of the report for the project.

The outcomes of the analyses of interview data and lesson observation data were triangulated with the teacher survey data to enhance the validity of the IPE data analysis. The process of triangulation involved the full-time researcher identifying and reviewing key themes from the teaching observations by comparing these with relevant data from interviews and with teacher survey responses to ensure that each potential finding was corroborated, or explained with more nuance, or rejected. These themes were reviewed by the PI to the project and ultimately by the whole research team as agreement was reached on final outcomes of the project. This final phase of analysis included the finalisation of main themes which ultimately served as a way of reporting the IPE findings in a way that commented, for example, on fidelity to the intervention and its effectiveness in the context of the impact evaluation outcomes and England's national curriculum.

5. Research ethics

The research was reviewed and cleared by the UCL Institute of Education's ethics committee, and conformed to the research ethics guidance of the British Educational Research Association (BERA).

6. Results

The writing tests results for the Y2 classes in 65 schools were successfully uploaded and analysed. Data needed for the primary analysis were collected for 1,246 pupils in total, with 637 in treatment and 609 in control (see Figure 1 for details of the allocation of participants). This led to an attrition rate of 28 percent of the total randomised sample. The control group had a higher attrition rate of 31.6 percent compared to the 24.6 attrition rate for the intervention group. The reasons for attrition were three school classes isolating due to Covid-19, and two school classes who missed the tests upload deadline which was prior to school summer holidays. Our quantitative analysis of the impact of attrition via observable factors did not substantially change the results of our study overall (see also the pre-registered statistical analysis plan – Anders et al., 2021). However, we regard the level of attrition in this study as a limitation for the confidence of our findings.

Tables 1 and 2 present the baseline characteristics of pupils and schools in the treatment and control groups. Table 1 reports this for the groups as randomised, while Table 2 does so for the groups as analysed. As such, changes in imbalance between the two tables are a result of sample attrition between the point of randomisation and collection of outcome measures.

At the point of post-COVID-19 randomisation, the control group had higher Sentence Generation (SG) test baseline scores (in baseline exercises carried out both pre- and post-COVID-19, where applicable) compared to the intervention group (most apparent pre-COVID-19). The control group also had higher Progress in English (PiE) baseline scores pre-COVID-19 compared to intervention groups. The intervention group had slightly higher PiE baseline scores than control schools post-COVID-19 (d=0.063) and contained higher proportions of pupils who had ever been eligible for free school meals (FSM; an indicator of coming from a low-income family), and pupils for whom English was an additional language (further information about the measures can be found in a section later in the paper).

At the point of analysis, the balance between control and intervention groups followed similar patterns to the balance at the point of randomisation. A sizeable imbalance in the pre-COVID-19 PiE baseline scores emerged in favour of the intervention group. Conversely, the balance for the PiE baseline score and the pre-COVID-19 SG test baseline scores improved, while the imbalance towards the treatment group in the proportion of pupils who had ever been eligible for FSM increased. Our pre-registered analytic approach anticipated the potential for imbalance and, consequently, sought to adjust for such baseline differences.

The results of the effects of the intervention on pupils' SG and on their writing of narrative text (PiE) can be seen in Table 3.

Table 1. *Baseline characteristics of groups as randomised*

	Control		Intervention		Overall			
Baseline variable	n (missing)	Mean/ Proportion (SD)	n (missing)	Mean/ Proportion (SD)	n (missing)	Mean/ Proportion (SD)	Standardised difference	p-value
PIE Short/ Baseline Score	794 (97)	5.76 (3.55)	774 (71)	5.98 (3.45)	1568 (168)	5.67 (3.50)	-0.063	.211
SG Baseline Score	759 (132)	14.72 (10.43)	764 (81)	13.81 (9.43)	1523 (213)	14.26 (9.95)	0.092	.077
Female	891 (0)	0.49 (0.50)	845 (0)	0.51 (0.50)	1736 (0)	0.50 (0.50)	-0.040	.316
Ever FSM	891 (0)	0.21 (0.41)	845 (0)	0.28 (0.45)	1736 (0)	0.24 (0.43)	-0.163	.000
English as an additional language	891 (0)	0.40 (0.49)	845 (0)	0.45 (0.50)	1736 (0)	0.42 (0.49)	-0.101	.031
Pre-COVID PIE Baseline	509 (382)	6.43 (2.08)	521 (324)	6.33 (1.57)	1030 (706)	6.38 (1.84)	0.054	.355
Pre-COVID SG Baseline	494 (397)	15.49 (3.80)	521 (324)	14.92 (5.25)	1015 (721)	15.20 (4.61)	0.124	.048

Notes. The standardised difference column reports difference in means between control and intervention groups divided by the pooled standard deviation of the overall sample. The p-value column is derived from cluster-robust t-tests of the null hypothesis of no difference in means between the treatment and control groups. Item missing values (in parentheses) report reduction in sample size for this variable relative to full sample as randomised. Ever FSM = Ever eligible for Free School Meals, an indicator of the participant coming from a low-income background.

Table 2. *Baseline characteristics of groups as analysed*

	Control		Intervention		Overall			
Baseline variable	n (missing)	Mean/ Proportion (SD)	n (missing)	Mean/ Proportion (SD)	n (missing)	Mean/ Proportion (SD)	Standardised difference	p-value
PIE Short/					1246 (322)	5.90 (3.50)		.429
Baseline Score	609 (185)	5.82 (3.59)	637 (137)	5.98 (3.42)			-0.046	
SG Baseline Score	609 (150)	15.26 (10.40)	637 (127)	14.07 (9.38)	1246 (277)	14.65 (9.91)	0.120	.035
Female	609 (282)	0.49 (0.50)	637 (208)	0.50 (0.50)	1246 (490)	0.50 (0.50)	-0.020	.648
Ever FSM	609 (282)	0.16 (0.36)	637 (208)	0.26 (0.44)	1246 (490)	0.21 (0.41)	-0.248	.000
English as an Additional					1246 (490)	0.42 (0.49)		.103
Lang.	609 (282)	0.39 (0.49)	637 (208)	0.44 (0.50)			-0.101	
Pre-COVID PIE Baseline	346 (163)	6.13 (1.89)	409 (112)	6.39 (1.56)	755 (275)	6.27 (1.72)	-0.151	.041
Pre-COVID SG Baseline	334 (160)	15.45 (3.97)	409 (112)	14.97 (5.13)	743 (272)	15.19 (4.65)	0.103	.164

Notes. The standardised difference column reports difference in means between control and intervention groups divided by the pooled standard deviation of the overall sample. The p-value column is derived from cluster-robust t-tests of the null hypothesis of no difference in means between the treatment and control groups. Item missing values (in parentheses) report reduction in sample size for this variable relative to non-missing variable values in sample as randomised. Ever FSM = Ever eligible for Free School Meals, an indicator of the participant coming from a low-income background.

Table 3. *Multilevel models of primary and secondary outcomes*

	Progress in English (PiE)	Sentence Generation (SG)
Treatment	0.122	1.620
	[-0.705,0.950]	[-0.676,3.916]
Baseline Progress in English score	0.434***	0.679***
	[0.334,0.534]	[0.480,0.879]
Baseline Sentence Combining score	0.103***	0.456***
	[0.063,0.143]	[0.371,0.542]
School Average Baseline Progress in English score	-0.039	-1.380**
	[-0.343,0.265]	[-2.377,-0.383]
School Average Baseline Sentence Combining score	-0.031	0.149
	[-0.139,0.076]	[-0.200,0.498]
Female	1.142***	1.198*
	[0.664,1.621]	[0.191,2.206]
Ever eligible for Free School Meals	-0.229	-0.647
	[-0.824,0.366]	[-1.551,0.256]
English as an Additional Language	0.177	-0.238
	[-0.343,0.697]	[-1.452,0.976]
School Prop. Female	-2.582	-1.835
	[-5.358,0.193]	[-12.310,8.640]
School Prop. Ever eligible for Free School Meals	2.839	-4.446
	[-0.960,6.638]	[-14.256,5.363]
School Prop. English as an Additional Language	0.141	-5.685
	[-2.153,2.435]	[-12.627,1.257]
Constant	3.434	15.484**
	[-0.822,7.690]	[5.065,25.904]
Cohen's d treatment effect size	0.026	0.145
Lower CI of Cohen's d	-0.152	-0.060
Upper CI of Cohen's d	0.205	0.350
ICC	0.153	0.245
Within R2	0.269	0.319
Between R2	0.529	0.456
Overall R2	0.324	0.352
School-level standard deviation	1.535	4.673
Pupil-level standard deviation	3.608	8.212
N pupils	1246	1273
N schools	63	63

Notes. Effect sizes are obtained from the coefficient on the treatment variable (β_1) in our analysis linear regression models, divided by the unconditional pooled standard deviation of the relevant outcome measure across treatment and control groups.

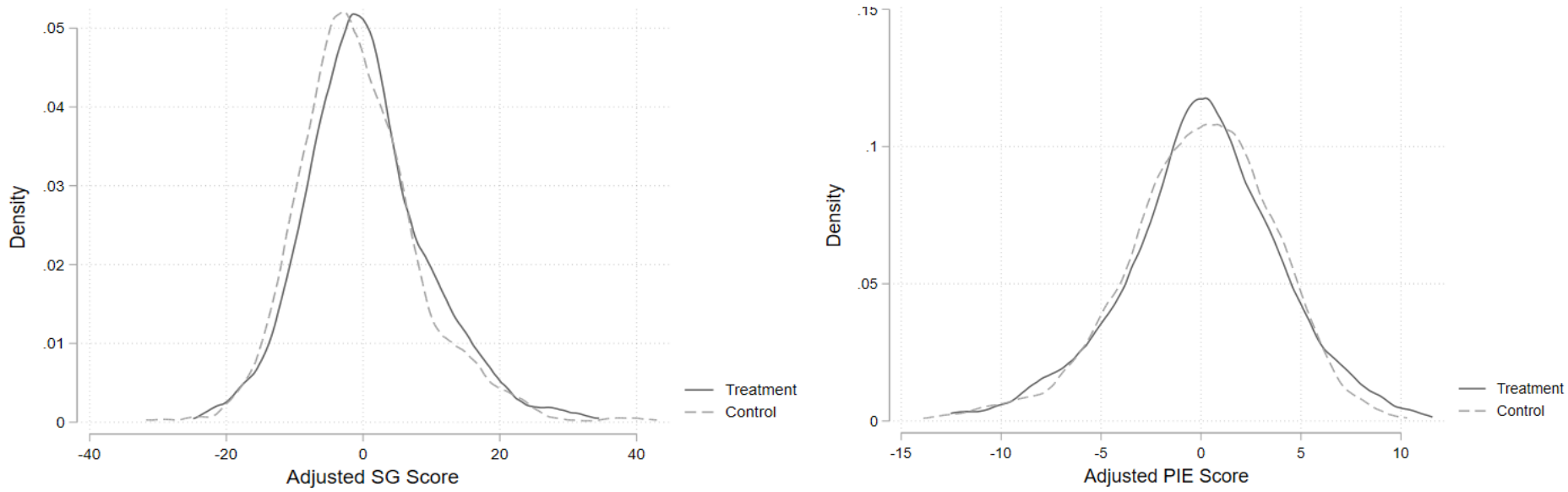


Figure 2. Distribution plot of outcome measures by treatment group[3]

[3] Notes. Outcome measures adjusted by using a linear regression model (as per pre-specified primary and secondary analysis models, with the exception of exclusion of the treatment variable) then calculating residuals from this model.

The unadjusted means for the primary outcome (PiE) were 9.78 for the control group and 9.36 for the intervention (i.e., slightly higher in the control group). However, after adjusting for covariates using the multilevel model (which, among other things, adjusts for the fact that the proportion of FSM pupils was higher in the treatment group), we estimate a very small Cohen's d effect size of 0.03. There was no statistically significant difference between the two groups (p = .77).

The unadjusted mean for the secondary outcome (SG) was 16.96 in the control group and 17.23 in the intervention group (i.e., slightly higher in the intervention group). After estimating our multilevel model including the baseline covariates (which, among other things, adjusts for the fact that the SG baseline score was higher in the control group) we estimate a moderate effect size of 0.14, however, there was still no statistically significant difference between the two groups (*p* = .17).

For further illustration of the data underlying our estimates, we plot the distribution of our outcome measures (after adjusting for baseline measures using the relevant analysis model) in Figure 2 (PiE in the left panel; SG in the right panel). This provides a visual representation of the differences in outcome measures between the treatment and control groups on which our impact estimates rest.

Noting the possibility that this intervention may have affected the distribution of writing attainment, we also conducted exploratory analysis of differences in variance between the treatment and control groups using variance ratio tests (e.g. Bland, 2015, pp. 144-145), albeit that these do not fully account for the clustered nature of our data so should be interpreted with caution. We find evidence of increased variance in the treatment group writing scores (p=0.02). To understand this further, we explored differences in within-group and between-group variances (again using variance ratio tests, but applied separately to within-group transformed data and group average data, respectively), finding that this overall increase appears to be driven by increased between-group variance (p=0.05) rather than within-group variance (p=0.37). This suggests that, rather than increasing within-class inequality, the intervention may have lifted scores in some schools but not others, although we stress the exploratory nature of this analysis and risks of false positive findings. Nevertheless, it suggests the possibility of school-level heterogeneity for future exploration. We find no evidence of differences in variance for sentence generation scores.

6.1 Sub-group and complier analyses

We additionally considered whether there was evidence of differential effects among two different sub-groups — pupils eligible for Free School Meals (FSM), pupils with English as an Additional Language (EAL)[4] — as well as among schools who fully complied with the intervention.

[4] We also explored whether we saw different impacts among Batch 3 schools (those randomised post-COVID 19) because these were least affected by COVID-19 disruption. We, hence, saw this as a check that our results were not influence by this disruption. The outcome of an analysis of Batch 3 data suggested that losing schools from the project due to COVID-19 disruption, ahead of the project restart, was not driving our findings. We also carried out robustness checks on Batch 1 and 2 schools using their

Table 4 reports the interaction effects (i.e., the difference between the effect for the sub-group and the effect in the rest of the sample), along with the number of pupils in each sub-group, and the Cohen's d effect sizes of the intervention among the sub-group estimated using multilevel models. For the primary outcome, the estimate on the interaction for FSM pupils was very small (d = -0.01) with a large p-value of 0.76, implying minimal difference between effects for FSM and non-FSM pupils. However, for EAL pupils there was a larger estimated interaction effect of d = –0.18 with a smaller (though still insignificant) p-value of 0.41.

The secondary outcome followed a similar pattern with a small statistically insignificant interaction estimate for FSM pupils (d = 0.08), implying little difference in effects between FSM and non-FSM pupils, but a larger negative interaction effect for EAL students (i.e., difference in effects of EAL and non-EAL pupils; d= -0.22) with a p-value of 0.12. This provides indicative evidence that the intervention was less effective for EAL students than it was for the cohort as a whole, but also suggests a more encouraging impact of the intervention on the SG score among non-EAL students (d = 0.24), albeit not statistically significant.

We knew from our contacts with schools that the main reasons for attrition were as follows: three teachers said that COVID-19, e.g. isolation of their class, was the problem; for two schools they were unable to submit a batch of final test outcomes due to errors in their processes and/or the school closed for summer holidays. However, it is also important to remember that our quantitative analysis of the impact of attrition via observable factors did not substantially change the results of our study overall. Nevertheless, we regard the high level of attrition as a limitation on the confidence in our findings.

In relation to fidelity of duration of lessons, results from the complier analysis do not explain the lack of impact on the primary outcome. The estimated effect size for those who complied fully with intervention delivery as defined is 0.05 (similar to the overall effect size of 0.03 – see Table 3). This effect was not statistically significant (p = 0.67). The estimated effect on the primary outcome of having done one additional standard deviation's worth of lessons was 0.02 and not statistically significant (p = 0.67). Accounting for the extent of time for implementation of the intervention thus did not change the overall impact evaluation results.

---

pre-COVID baseline measures as covariates instead of those collected at point of project restart. Again, this did not make any difference to our findings suggesting that this initial school loss is not driving our main findings.

Table 4. *Multilevel model of primary and secondary outcomes in sub-groups*

	Progress in English (FSM)	Sentence Generation (FSM)	Progress in English (EAL)	Sentence Generation (EAL)
Treatment	0.073	1.723	0.395	2.525
	[-0.865,1.011]	[-0.699,4.144]	[-0.528,1.317]	[-0.017,5.068]
Treatment X FSM	0.252	-0.491		
	[-0.938,1.443]	[-2.465,1.483]		
Treatment X EAL			-0.644	-2.125
			[-1.552,0.263]	[-4.368,0.119]
Baseline Progress in Writing score	0.433***	0.680***	0.436***	0.686***
	[0.334,0.533]	[0.479,0.882]	[0.337,0.534]	[0.484,0.887]
Baseline Sentence Combining score	0.103***	0.457***	0.103***	0.454***
	[0.063,0.143]	[0.372,0.542]	[0.063,0.143]	[0.368,0.541]
School Average Baseline Progress in Writing score	-0.039	-1.385**	-0.040	-1.373**
	[-0.343,0.266]	[-2.386,-0.384]	[-0.343,0.263]	[-2.372,-0.373]
School Average Baseline Sentence Combining score	-0.033	0.152	-0.028	0.158
	[-0.142,0.076]	[-0.201,0.505]	[-0.133,0.078]	[-0.197,0.513]
Female	1.142***	1.199*	1.137***	1.180*
	[0.664,1.620]	[0.188,2.210]	[0.662,1.612]	[0.170,2.190]
Eligible for Free School Meals	-0.380	-0.352	-0.220	-0.625
	[-1.302,0.541]	[-1.891,1.187]	[-0.811,0.371]	[-1.532,0.282]
English as an Additional Language	0.173	-0.231	0.521	0.960
	[-0.344,0.690]	[-1.447,0.985]	[-0.043,1.086]	[-0.309,2.229]
School Prop. Female	-2.546	-1.982	-2.654	-2.035
	[-5.367,0.274]	[-12.488,8.524]	[-5.391,0.083]	[-12.623,8.552]
School Prop. Free School Meals	2.803	-4.374	2.890	-4.390
	[-1.001,6.606]	[-14.213,5.465]	[-0.884,6.664]	[-14.345,5.564]
School Prop. English as an Additional Language	0.147	-5.674	0.109	-5.953
	[-2.149,2.443]	[-12.613,1.266]	[-2.165,2.383]	[-12.814,0.909]
Constant	3.488	15.421**	3.307	15.150**
	[-0.824,7.801]	[4.863,25.979]	[-0.960,7.574]	[4.730,25.571]
Sub-Group Cohen's d	0.076	0.120	-0.057	0.036
Lower CI of Cohen's d	-0.164	-0.122	-0.273	-0.197
Upper CI of Cohen's d	0.316	0.361	0.159	0.268
ICC	0.149	0.217	0.156	0.250
Within R2	0.270	0.319	0.270	0.321
Between R2	0.528	0.454	0.535	0.456

Overall R2	0.324	0.352	0.326	0.355
School-level standard deviation	1.509	4.330	1.548	4.740
Pupil-level standard deviation	3.609	8.214	3.607	8.202
N pupils	1246	1273	1246	1273
N schools	63	63	63	63
N pupils in subgroup	259	274	518	539

Notes. See Table 3 for description of calculation of Cohen's d calculation. Cohen's d within sub-group derived from addition of coefficient on treatment variable and coefficient on interaction of treatment variable and sub-group variable (β_1+β_3). FSM = Eligible for Free School Meals; EAL = Has English as an Additional Language.

6.2 Qualitative Implementation and Process Evaluation

The findings from the IPE included questionnaire survey and lesson observation data. Tables 5 shows information about the experiences of primary school teaching reported by respondents in the pre-intervention survey (which was distributed to all teachers in the research).

Table 5. *Teachers' Experience Teaching in Schools*

Teaching Experience	Number of Teachers
Teaching in current and other schools	
0–2 years FTE	6
3–5 years FTE	16
6–10 years FTE	13
More than 10 years FTE	11

Primary Year Group Taught	Number of Teachers
Year 1 (Age 5–6)	22
Year 2 (Age 6–7)	45
Year 3 (Age 7–8)	14
Year 4 (Age 8–9)	14
Year 5 (Age 9–10)	15
Year 6 (Age 10–11)	6

Table 5 shows that the participant teachers were mostly experienced teachers, with 40 out of 46 (87%) having at least three years of teaching experience.

The pre intervention survey provided some general patterns across the whole sample of teachers prior to random allocation to groups and the experience of intervention or control conditions. Before being part of the grammar intervention the frequency of the use of approaches that the teachers reported used to teach writing, including the teaching of grammar, varied. Table 6 shows the responses related to approaches to supporting the composition of writing including attention to grammar.

The majority of teachers who responded to this survey question reported drawing pupils' attention to grammatical structures at least weekly as part of their teaching of writing, although the seven teachers who reported never doing this was an unexpected finding in view of the requirements of England's national curriculum which requires explicit grammar teaching. A relatively large number of teachers reported never giving pupils choice over topic, form and audience for writing, although a majority did report offering choice over writing topics and some other aspects of writing. These data showed overall the notable variation in emphases in approaches to grammar and writing teaching prior to the intervention.

Table 6. *Response to Survey Question 11: Supporting the composition of writing*

Q11. Which of the following approaches do you use, if any, to support composition of writing with your current class?					
	never	hourly	daily	weekly	monthly
Set activities that require pupils to vary the formality of written language (e.g., letters to friends v. report writing).	2	0	8	24	13
Teach pupils to make choices in relation to topics and ideas for their writing	3	0	17	20	7
Draw pupils' attention to differences in meaning between specific grammatical structures.	7	1	9	23	7
Teach pupils to analyse the forms of texts they read as a stimulus for their writing	4	0	16	20	7
Encourage pupils to choose topic, form, and audience for writing, and decisions on when the writing is finished.	13	1	7	18	8

6.3 Teachers' views about teaching the national curriculum grammar requirements

During the first interviews of teachers, as part of the visits to schools, when asked an explicit question about the national curriculum teachers 060, 010, 033, and 003 described the grammar requirements in similar ways, e.g., using phrases about the grammar content such as 'too much' and 'a lot'. A fifth teacher felt the specification pushed pupils too early in the year and was 'very idealistic' (014).

When the teachers in the intervention group were surveyed the majority felt that knowing technical terms was essential for learning to write, and that the national curriculum grammar requirements were a positive feature. However, as can be seen in Table 7, 11 out of the 16 teachers who responded also said that their views about the national curriculum had been influenced by their involvement in the *Englicious* intervention (answer to question 5d).

Table 7. *Answers to the survey questions that showed views about grammar teaching and the national curriculum in England (mid-intervention survey).*

Questions	Strongly Agree	Agree	Disagree	Strongly disagree
Q5a. Knowing technical terms for grammar (such as noun, phrase or clause) is essential for learning to write.	3	10	4	0
Q5b The national curriculum's requirements for teaching grammatical terms in Y2 are a positive feature that helps teachers to improve children's writing.	2	13	2	0
Q5d To what extent have your views been influenced by your involvement in the *Englicious* intervention?	Significantly influenced	Somewhat influenced	Not influenced much	Not influenced at all
	2	9	5	0

These frequencies of types of responses were very similar to the frequencies in the end of intervention survey which attracted fewer responses in total (*n* = 11).

Survey data collected at the mid-point during the intervention period, from 19 teachers in the intervention group (out of a total of 33 teachers), indicated that the majority of respondents agreed that the lessons were having a moderate positive effect on the way that pupils were writing. The end-of-intervention survey also showed 10 out of the 11 teachers who responded agreed that the intervention lessons had had a positive effect on pupils' writing. In terms of the relatively small number of critical observations made about *Englicious*, concerns included a perception that some of the IWB/PowerPoint slides contained too much text; the amount of content to be covered in lessons impacting on what was perceived by some to be the slow pace of lessons; and perceptions that the lessons were relatively passive for the pupils.

As we have commented above, overall there was a high level of fidelity to the intervention. Of the 32 observed and noted cases of deviations from the intervention requirements only four instances were considered significant. Three of these had involved more time spent on the practice of writing at the end part of the *Englicious* lesson, and one deviation was the use of additional scaffolds for the pupils writing in this final part of the lesson.

One key difference between intervention and control schools, found as a result of the lesson observations, was that the *Englicious* lessons consistently included an opportunity for pupils to apply their learning through an independent writing activity that was part of the *Englicious* lesson. This was not a typical approach in every lesson observed in the control

schools. In the control schools a wide range of teaching strategies was seen being used to support learning about grammar, for example general approaches to grammar teaching that included using a text to contextualise teaching of grammatical terms and their properties; teacher-led strategies including deliberate inclusion of errors when presenting texts; whole-class activities including discussions while pupils were sitting on the carpet; and use of mini whiteboards for pupils to write sentences, and other examples which also allowed for the teachers to formatively assess pupils.

7. Discussion

Using a RCT and IPE we examined the impact of a new grammar teaching intervention designed to support England's national curriculum requirements for grammar and writing for six-year-old and seven-year-old children. Children's writing was measured at text-level and sentence-level. There was no significant effect of the intervention on our primary outcome measure, the test of narrative writing. The lack of impact on the primary outcome is in line with the results of previous studies and meta-analyses of primary age children (cited at the beginning of this paper), although there was no evidence of an overall negative effect of explicitly teaching grammar which has been found in some studies. By contrast, in children's performance on the SG test there was a positive, but non-significant, effect of the intervention (effect size 0.14). The non-significant outcome means that we cannot rule out factors other than the intervention as contributors. Our interpretation of the effect size measurements is informed by recent trends of thinking about statistical analyses generally but also in education interventions (Kraft, 2020; Wasserstein et al., 2019). As such we tentatively view the effect size for pupils' sentence generation, while acknowledging the uncertainty in our estimate, as worthy of further exploration since it is not one that has been found in previous research of grammar for writing approaches with children aged six to seven-years-old.

The IPE revealed that although the teachers' practices for teaching grammar and writing in the business-as-usual control condition were to some extent following the national curriculum requirements for grammar, their approaches were much more varied than in the intervention condition. The intervention condition was explicitly designed, by the linguists on the team, and teachers, to meet the requirements of the grammar content of the national curriculum. The teachers in the intervention group carried out this teaching with high levels of fidelity. Nearly all the case study teachers thought that *Englicious* had a positive effect. The teachers' views about the national curriculum requirements for grammar were varied, and where positive appeared to have been influenced by their involvement in the intervention.

7.1 Limitations of the research

It is important to acknowledge limitations in the research and its design that necessitate further research in future. We cannot exclude the possibility that any effects of the intervention were due to chance. The effect size of 0.03 for the primary outcome has a 95% confidence interval, from -0.15 to 0.21, meaning that we cannot exclude the possibility that the intervention had a negative effect of about a sixth of a standard deviation or a positive

effect of nearly a third of a standard deviation at this level of statistical significance. Similarly, the secondary outcome has an effect size of 0.14 which has a 95% confidence interval from -0.06 to 0.35, meaning that we cannot exclude a negative effect of about a twentieth of a standard deviation or a positive effect of about a third of a standard deviation at this level of statistical significance. With regard to the sentence generation task, as the same word pairs were used at baseline and immediate post intervention it is possible that there could have been gains due to the children remembering the task although the tests were implemented approximately three months apart.

We also highlight the important context for business as usual in our setting. The pupils in the business-as-usual control group did experience some grammar teaching but this was not explicitly connected with practising writing in the way that it was in the *Englicious* intervention group. The intervention group had practice in writing, something that is not consistently done in regular teaching (Dockrell et al., 2015).

Although despite the effects of COVID-19 lockdowns we were able to work with 70 primary schools we also acknowledge that the attrition rate of schools in the study was relatively high in relation to our impact analyses. Attrition is regarded as a potential cause of bias in a RCT because it may introduce selection bias between treatment and control groups in terms of factors that were not measured in the study. Such factors could be an alternative explanation for the results, although we have done our best to check for such effects (e.g., through our analysis of only Batch 3 schools). Finally, there remains a challenge in devising reliable and valid assessments of young children's narrative writing in early elementary/primary school to capture variation in performance.

8. Conclusions

Based on the outcomes of our research we hypothesise that two factors could potentially explain the changes in pupils' sentence-generation test outcomes: 1. The requirement for pupils to experience, discuss and think about the direct manipulation of words, phrases and sentences in lessons; 2. The direct connections made between grammar teaching and practising writing. These two factors were consistently applied and connected as part of the intervention group teaching but not the control group who had much more varied types of grammar teaching including more examples where grammar was less well connected to the practice of writing. Practice in writing is an important contributor to writing development generally, and children in the intervention group had more practice, so the driver of the small positive effect on sentence generation remains a hypothesis. What's more, we are uncertain about why there was no discernible effect on the pupils' narrative writing.

New research is needed to evaluate how manipulation of words, phrases and sentences might be combined within lessons to enhance young children's writing skills. Evidence-based approaches to teaching writing that have shown gains for pupils' narrative writing could also be drawn upon for new interventions. There also remains the need to learn more about which particular kinds of grammar knowledge are most useful to aid learners' improvement in use of written language.

The evidence reviewed and presented in this paper prompts some questions about whether the type of grammar content, and the amount of grammar content, in England's national curriculum for primary schools is appropriate. Some have argued that grammar has conceptual importance in its own right, and does not necessarily only need to be evaluated in relation to the teaching of writing (Aarts, 2019; van Rijt & Coppen, 2021). However, given the importance of young children learning to write, a focus on how children can master the grammar of sentences is a paramount concern. England's national curriculum has included relatively extensive grammatical knowledge of a particular kind, but there are other types of grammatical knowledge that could be specified in future revisions.

The research evidence from the field of studies on grammar and writing for children in elementary/primary education provides a robust basis for *some* decisions about content and pedagogy in national curriculum programmes of study for teaching writing. In view of the limited impact of grammar teaching on narrative writing with young children, shown in previously published studies and in our research, and the evidence that other approaches to teaching writing do have a beneficial impact on pupils' writing we conclude that the programmes of study for teaching writing in the national curriculum in England could usefully be reviewed in order to ensure a better fit with research evidence of what works in the teaching of writing.

Data availability

The anonymised data supporting the findings of this study are publicly available in UCL's Research Data Repository at https://doi.org/10.5522/04/30258862

Funding

We are grateful to the Nuffield Foundation for the two research grants that enabled us to carry out the research project reported in this paper.

References

Aarts, B. (2011). *Oxford Modern English Grammar*. Oxford University Press.

Aarts, B. (2019). *Teaching English grammar: the Englicious approach.* https://www.meits.org/dialogues/article/teaching-english-grammar-the-*englicious*-approach

Aarts, B., Mehl, S., & Wallis, S. (2016). Language learning at your fingertips. In K. Corrigan & A. Mearns (Eds.), *Creating & Digitizing Language Corpora* (pp. 211-239). Palgrave.

Aarts, B., & Smith-Dennis, E. (2018). Using corpora for English language teaching and learning. In D. McIntyre & H. Price (Eds.), *Applying linguistics: language and the impact agenda*. Routledge.

Abbott, R. D., & Berninger, V. (1993). Structural Equation Modeling Of Relationships Among Developmental Skills And Writing Skills In Primary-Grade And Intermediate-Grade Writers. *Journal of Experimental Psychology*, *85*(3), 478-508. https://doi.org/10.1037/0022-0663.85.3.478

Anders, J., Dockrell, J., Sing, S., Torgerson, C., & Wyse, D. (2021). *Grammar and Writing Research Project: Statistical Analysis Plan*. https://discovery.ucl.ac.uk/id/eprint/10129724/1/*Englicious*%20SAP%20v1.0.pdf

Anders, J., Wyse, D., Aarts, B., Dockrell, J., Torgerson, C., Manyukhina, Y., Cushing, I., & Sing, S. (2019). *Evaluation protocol: Grammar and Writing research project.* https://discovery.ucl.ac.uk/id/eprint/10087427/1/*Englicious*%20Evaluation%20Protocol%20v1%20FINAL.pdf

Andrews, R., Torgerson, C., Beverton, S., Freeman, A., Locke, T., Low, G., Robinson, A., & Zhu, D. (2004). *The effect of grammar teaching (sentence combining) in English on 5 to 16 year olds' accuracy and quality in written composition*. http://eppi.ioe.ac.uk/cms/

Arfé, B., Dockrell, J., & Bernardi, B. (2016). The effect of language specific factors on early written composition: the role of spelling, oral language and text generation skills in a shallow orthography. *Reading and Writing*, *29*, 501-527. https://doi.org/10.1007/s11145-015-9617-5

Atkinson, P., & Delamont, S. (2005). Analytic Perspectives. In N. Denzin, K. & Y. Lincoln, S. (Eds.), *The Sage Handbook of Qualitative Research* (Third ed., pp. 821-840). Sage.

Berninger, V., Nagy, W., & Beers, S. (2011). Child writers' construction and reconstruction of single sentences and construction of multi-sentence texts: contributions of syntax and transcription to translation. *Reading and Writing*, *24*, 151-182. https://doi.org/10.1007/s11145-010-9262-y

Bland, M. (2015). *An Introduction to Medical Statistics*. Oxford University Press.

Department for Education (DfE). (2013). *The national curriculum in England: Key stages 1 and 2 Framework document. September 2013.* Department for Education.

Dockrell, J., Connelly, V., & Arfé, B. (2019). Struggling writers in elementary school: Capturing drivers of performance. *Learning and Instruction*, *60*, 75-84. https://doi.org/10.1016/j.learninstruc.2018.11.009

Dockrell, J., Marshall, C., & Wyse, D. (2015). Teachers' reported practices for teaching writing in England. *Reading and Writing*, 409-434. https://doi.org/10.1007/s11145-015-9605-9

Drijbooms, E., Groen, M., & Verhoeven, L. (2017). How executive functions predict development in syntactic complexity of narrative writing in the upper elementary grades. *Reading and Writing*, *30*(1), 209-231. https://doi.org/10.1007/s11145-016-9670-8

Englicious Survey of English Usage. (2024). *Englicious Survey of English Usage*. http://www.*englicious*.org

Fontich, X., Van Rijt, J., & Gauvin, I. (2020). Intro to the Special Issue Research on L1 grammar in schooling: Mediation at the heart of learning grammar. *L1-Educational Studies in Language and Literature*, *20*, 1-13. https://doi.org/10.17239/L1ESLL-2020.20.03.01

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, *141*(1), 2-18. https://doi.org/https://doi.org/10.1037/a0024338

Gerber, A. S., & Green, D. P. (2012). *Field Experiments: Design, Analysis, and Interpretation*. W. W. Norton & Company.

Graham, S., Kim, Y.-S., Cao, Y., Lee, J. W., Tate, T., Collins, P., Cho, M., Moon, Y., Chung, H., & Olson, C. B. (2023). A meta-analysis of writing treatments for students in grades 6–12. *Journal of Educational Psychology*, *115*(7), 1004–1027. https://doi.org/https://doi.org/10.1037/edu0000819

Graham, S., McKeown, D., Kiuhara, S., & Harris, K. (2012). A Meta-Analysis of Writing Instruction for Students in the Elementary Grades. *Journal of Educational Psychology*, *104*(4), 879-896.

Graham, S., & Perin, D. (2007a). A Meta-Analysis of Writing Instruction for Adolescent Students. *Journal of Educational Psychology*, *99*(3), 445-476.

Graham, S., & Perin, D. (2007b). What We Know, What We Still Need to Know: Teaching Adolescents to Write. *Scientific Studies of Reading*, *11*, 313-335. https://doi.org/10.1080/10888430701530664

Imbens, G. M., & Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction.* Cambridge University Press.

Kawulich, B. (2017). Coding and Analyzing Qualitative Data. In D. Wyse, N. Selwyn, E. Smith, & L. Suter (Eds.), *The BERA SAGE Handbook of Educational Research* (pp. 769-791). SAGE.

Kim, Y. S. G., & Schatschneider, C. (2017). Expanding the Developmental Models of Writing: A Direct and Indirect Effects Model of Developmental Writing (DIEW). *Journal of Educational Psychology*, *109*(1), 35-50.

Kirkup, C., Reardon, T., & Sainsbury, M. (2006). *Progress in English 7: Teacher's Guide*. GL Assessment.

Koster, M., Tribushinina, E., de Jong, P. F., & van den Bergh, H. (2015). Teaching children to write: A meta-analysis of writing intervention research. *Journal of Writing Research*, *7*. https://doi.org/https://doi.org/10.17239/jowr-2015.07.02.2

Kraft, M. (2020). Interpreting Effect Sizes of Education Interventions. *Educational Researcher*, *49*(4), 241-253.

Myhill, D. (2021). Grammar re-imagined: foregrounding understanding of language choice in writing. *English in Education*, *55*(3), 265-278. https://doi.org/10.1080/04250494.2021.1885975

Myhill, D., Jones, S., Lines, H., & Watson, A. (2011). Re-thinking grammar: the impact of embedded grammar teaching on students' writing and students' metalinguistic understanding. *Research Papers in Education*, *27*(2), 139-166.

Torgerson, C., Torgerson, D., Mitchell, N., Buckley, H., Ainsworth, H., Heaps, C., & Jefferson, L. (2014). *Grammar for Writing: Evaluation Report and Executive Summary*. Education Endowment Foundation.

van Rijt, J., & Coppen, P.-A. (2021). Linguistic concepts in L1 grammar education: A systematic literature review. *Linguistics and Education*, *61*, 100890. https://doi.org/https://doi.org/10.1016/j.linged.2020.100890

Walter, K., Dockrell, J., & Connelly, V. (2021). A sentence-combining intervention for struggling writers: response to intervention. *Reading and Writing*, *34*, 1825-1850.

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond 'p < 0.05'. *The American Statistician*, *73*(sup1), 1–19. https://doi.org/https://doi.org/10.1080/00031305.2019.1583913

Wyse, D. (2001). Grammar. For Writing?: A Critical Review of Empirical Evidence. *British Journal of Educational Studies*, *49 (4)*, 411-427.

Wyse, D. (2006). Pupils' Word Choices and the Teaching of Grammar. *Cambridge Journal of Education*, *36*(1), 31-47.

Wyse, D. (2017). *How Writing Works: From the Invention of the Alphabet to the Rise of Social Media*. Cambridge University Press.

Wyse, D., Aarts, B., Anders, J., de Gennaro, A., Dockrell, J., Manyukhina, Y., Sing, S., & Torgerson, C. (2022). *Grammar and Writing in England's National Curriculum. A Randomised Controlled Trial and Implementation and Process Evaluation of Englicious.* https://discovery.ucl.ac.uk/id/eprint/10144257/

Wyse, D., & Hacking, C. (2024). *The Balancing Act: An Evidence-Based Approach to Teaching Phonics, Reading and Writing.* Routledge.

Wyse, D., & Torgerson, C. (2017). Experimental trials and 'what works?' in education: The case of grammar for writing. *British Educational Research Journal*, *43*(6), 1019-1047. https://doi.org/10.1002/berj.3315

Appendix: An example from the teaching manual of one of the intervention lessons

Lesson 1: Nouns

Lesson objective: grammar

To understand that nouns are words that represent people, places and things, to understand how they are formed, and to understand what some of their grammatical characteristics are.

Lesson objective: writing

To explore what role nouns might play in writing, and to apply this to pupils' own writing.

Relevant National Curriculum grammatical terms

noun, singular, plural, possessive, suffix

Estimated time in minutes	Activity	Pedagogical rationale
10	**Starter** The teacher shows pupils labelled images that represent different nouns (e.g. *London*, *rabbit*, *Harry*). **Question for whole class discussion and responses:** What kinds of things do these words represent? The teacher explains that they all represent **nouns** and provides a working definition: nouns represent people, places and things. This definition is exemplified by returning to the images and talking through each of these: e.g. a *rabbit* is a noun because it represents a living thing; *London* is a noun because it represents a place; *Billy* is a noun because it represents a person. **Questions for whole class discussion and responses:** • What other nouns can you think of? • How do we know when a word is a noun? • Why is *happy* not a noun? If pupils provide examples, the teacher asks them to explain whether they represent people, places or things. Corrections are given by the teacher if these are not nouns. In these cases, the teacher might invite the class to discuss *why* words are not nouns. There is no requirement or	This activity is designed to get pupils thinking about what constitutes a 'noun', providing some initial examples to do so.

reason to explain what categories the non-noun words fall into.

10 **Teach**

Teacher-led explanation of the grammatical properties of nouns, using the following checklist. Examples should be provided, and pupils should be invited to offer their own examples.

This activity consolidates pupils' understanding of nouns.

- **Nouns can name one thing, person, animal, etc. or more than one thing**. When a noun names one thing we call it a **singular noun** (e.g. *table*). When it names more than one thing we call it a **plural noun** (e.g. *tables*). To make a noun into a plural, we normally add an *–s* on the end. Can you think of another example? Can you think of some exceptions?
- **Nouns can come after a word such as *the*, *a*, *an*, *some*, and *your*.** These words identify the noun that comes after them. For example, *the weekend*; *a dream*; *your pencil case*; *some fruit*. Can you think of another example?
- **Nouns can show belonging**. These are called **possessive nouns**. For example, the noun *Billy* can be changed to *Billy's* as in *Billy's coat*. Possessive nouns usually take an *–'s* ending. Can you think of another example?

Here, the teacher should also explain that some nouns can't be seen or touched. For example, *dream*, *love*, *mystery*. But these are still 'things', and have the 3 properties as described above. So we still want to call them nouns! Can pupils think of anymore?

10 **Practice**

Led by the teacher using an interactive whiteboard, the whole class carry out the 'Noun identification' activity on *Englicious*: selecting words from a list and dragging them into one of three categories: (1) noun; (2) not noun, or (3)

This activity consolidates pupils' understanding of nouns and

'?'. In the category '?' you can place words which you are not sure about.

This should be a lively and interactive activity: pupils might make use of the interactive whiteboard, for example, and teachers should invite pupils to justify and explain their decisions why they think words are nouns or not nouns, using what they have learnt about the 3 grammatical characteristics of nouns.

asks them to further explain and justify their decisions for considering whether a word is a noun or not.

How much time should you spend on this activity? For each lesson, make sure that before you teach it you give yourself a good idea of how much time you have to spend on an activity. If you have plenty of time you can spend more on it than if you don't, but in the latter case please don't skip the activity. See the answer to the 5th Frequently Asked Question below for further guidance.

20 **Apply**

In the final part of the lesson, the attention turns to writing.

First, the teacher shows three extracts from texts, shown below, and asks pupils to identify the nouns. If pupils struggle to do this, or suggest words that are not nouns, the teacher should run through the checklist for nouns with these words.

This activity asks pupils to apply their knowledge of nouns to writing: first by exploring the use of nouns in an existing fictional text, and then by creating their own.

Taller than a house, the Iron Man stood on the cliff, on the very edge, in the darkness.

and

Passengers travel down to underground stations in escalators. They speedily travel around the city in trains, through tunnels deep under the ground.

And then a final one, which pupils then explore some questions about:
It was Monday, it was pouring with rain, and it was the first day back at school after the holidays. That's why the class were in a bad temper.

Questions for whole class discussion and responses:

- Which word tells us know what day it was?
- What was the weather like, and how do you know? How were the pupils feeling, and how do you know?
- Do the nouns tells us about people, places or things?

The discussion should lead towards the idea that nouns can play an important role in telling readers about people, time and places.

Now the class think about how the writing could have been different, again in a series of exploratory questions:

- How could some of these nouns be given extra information? For example: *it was pouring with <u>cold rain</u>*. Can you change some of the nouns?
- Choosing one example, which do you prefer? Why?
- Why might a writer want to add extra detail to nouns?

Pupils then write a short text of their own. This should be the opening to a story. They should use nouns in order to fill the story with people, places and things.

Some guidance could be given here in terms of topic: i.e. it could be about a first day back at school following the holidays (as in the example text), or it could be about the first day of the summer holidays.

Pupils should then:

- Use nouns in order to fill the story with people, places and things.
- Circle the nouns in their writing.
- Share their writing with a partner, and discuss the nouns used in each other's writing, and what they do in the story.

The discussion should lead towards the idea that nouns can play an important role in establishing people, places, ideas and things in stories. They help to build and 'populate' fictional story worlds.

5	The teacher asks the pupils to think of three things they have learnt about nouns. Pupils think to themselves, share with each other and then the class.	This consolidates and re-caps information from the lesson.