Rehearsal Room Writing – Efficacy Trial
Evaluation Protocol

National Foundation for Educational Research (NFER)
Helen Poet

Template last updated: December 2022

Evaluation summary

Project title	Rehearsal Room Writing – Efficacy Trial
Developer *(Institution)*	The Royal Shakespeare Company (RSC)
Evaluator *(Institution)*	National Foundation for Educational Research (NFER)
Principal investigator(s)	Helen Poet
Protocol author(s)	Lillian Flemons, Ruth Staunton, Professor Philip Durrant, Helen Poet, Professor Annabel Watson
Trial design	Two-arm cluster randomised controlled trial with random allocation at school level
Trial type	Efficacy
Pupil age range and Key stage	Year 5
Number of schools *(at design stage)*	183
Number of pupils *(at design stage)*	5,600
Primary outcome measure and source	Writing Assessment Measure Comparative Judgement
Secondary outcome measure and source	Lexical diversity, lexical sophistication, dependency grammar framework, Liking Writing Scale (LSW), Self-efficacy for Writing Scale (SEWS), Teacher Efficacy Scale for Writing, KS2 writing, KS2 reading, KS2 maths.

Protocol version history

Version	Date	Reason for revision
1.0 [*original*]	10th September 2025	N/A
1.1	13th February 2026	Statistical methods expanded and updated to match methodology in the SAP, specifically RQ10 now has a binary outcome, approach to baseline variables for RQ11 and RQ12 has been refined, and class has been added as a model covariate for RQ1, RQ2, RQ4, RQ5, RQ6, RQ9, RQ10, RQ11 and RQ12. More detail about the judging process and safeguarding processes added.

Contents

Tables

Figures

Study rationale and background

In 2023/2024, only 72% of state-school pupils in England met the Key Stage 2 (KS2) standards in teacher-assessed writing (Department for Education, 2025c). This is a lower proportion than for reading, maths and science, and represents a 6 percentage point decrease from the upward trend seen prior to the COVID-19 pandemic (Department for Education, 2025c). Only 58% of disadvantaged pupils met the expected standard for teacher-assessed writing for KS2, compared to 78% of pupils not known to be disadvantaged (Department for Education, 2025a). Along with mathematics, this is the largest gap of any subject. Recent developments include the publication of an evidence-based writing framework (Department for Education, 2025d) and writing featuring as one of the core subjects for the 2024-25 Curriculum and Assessment Review (Department for Education, 2024a). Evidence from the National Literacy Trust indicates that children's writing enjoyment levels are 'at an unprecedented low', with fewer than half (47%) of children aged 8 to 11 years reporting to enjoy writing in their free time (Clark, Picton and Cole, 2024). This represents a drop of 6 percentage points over four years. The Education Policy Institute (EPI) has identified a lack of evidence around best practice for teaching writing as a key challenge for improving writing outcomes (McGlade and Bergin, 2025).

An evidence review of writing approaches by the Education Endowment Foundation (EEF) found good writing outcomes to be associated with programmes that attempt to build students' motivation to write and enjoyment of self-expression (Slavin *et al.*, 2019), reflecting a link between motivation to write and writing outcomes (Pajares, 2003; Troia *et al.*, 2013; Graham *et al.*, 2017; Grima *et al.*, 2024). Existing evidence also suggests a positive relationship between enjoyment of writing and writing achievement (Zumbrunn *et al.*, 2019).

There is some evidence of drama-based interventions having a positive impact for writing outcomes (Laurin, 2010; Education Endowment Foundation, 2021). Moreover, these effects appear to be strongest when the intervention sessions are led by a classroom teacher or researcher (rather than a performance arts professional), more than five such lessons take place and when they are integrated into the English language, arts or science curriculum (as compared to other domains) (Lee *et al.*, 2015). Collaborative learning and other social interactions have also been found to facilitate student motivation and learning, both generally and in relation to writing specifically (Camacho, Alves and Boscolo, 2021). There is also evidence to suggest that 'embodied learning' improves academic performance and reduces cognitive load by increasing active pupil engagement, and thereby deepening understanding and retention of knowledge (Lyu and Deng, 2024).[1]

Finally, existing research emphasises the importance of in-depth professional development around teaching writing, with teachers experiencing themselves the writing strategies that their

[1] Embodied learning refers to the use of bodily experience, sensory input and personal feeling (via playing parts, moving around a space, working with others and being exposed to theatrical stimuli) for constructing meaning and interpreting texts.

pupils would then employ (Slavin *et al.*, 2019) and having agency to tailor writing practices to meet the needs of their class (Education Standards Research team, 2012).

Rehearsal Room Writing (RRW) is a programme that aims to improve Year 5 pupils' writing ability by increasing their skills, motivation, self-efficacy and enjoyment of creative writing through drama-based approaches to engaging with Shakespeare's texts. The programme was developed by the Royal Shakespeare Company (RSC) as part of their training offer for primary school teachers. Feedback from similar work with 280 state schools and 15 regional theatres across England led the RSC to consider the programme's potential for impact in relation to writing outcomes. While the RRW programme is not currently commercially available to schools, the RSC offers professional development and resources for teachers using some of the Rehearsal Room techniques that are implemented as part of the programme.

While not designed as a writing intervention, existing evidence described above supports the idea that mechanisms within the RRW programme have the potential to positively impact pupils' writing ability. For example, one of the key objectives of the programme is to increase pupil enjoyment of and, consequently, motivation to write.

In the 2022-23 academic year, the RSC carried out an exploratory randomised control trial (RCT) of Year 5 pupils and teachers in 45 state-maintained primary schools. The trial was funded by the Paul Hamlyn Foundation and investigated the impact of Rehearsal Room teaching approaches in schools on a proximal writing measure[2] as well as the LanguageScreen Extended, ReadingScreen and Myself as a Learner Scale (McCulloch and Collins, 2023). While the teaching approaches used in the RRW programme have been used with a range of different age groups, Year 5 was selected as the focus for the trial as teacher feedback suggested that they found this to be the developmental stage where pupils most successfully channelled the active learning they experienced into their writing outputs. The programme was also seen to align with the emphasis placed by the National Curriculum for Key Stage 2 on narrative, description and cohesion as aspects of composition. While Year 6 would also have been appropriate in these respects, competing priorities around SATs and the transition to secondary settings resulted in a preference for Year 5.

The study reported positive outcomes for writing ability in terms of sophistication of vocabulary and complexity of sentence structures, using the Lexical Complexity Analyzer and TASSC, respectively, for pupils in intervention schools as compared to control schools (McCulloch and Collins, 2023). Positive outcomes were also observed in relation to pupils' self-concept as a learner, measured using the Myself as a Learner Scale (MALS) (McCulloch and Collins, 2023). Insufficient data was collected for the LanguageScreen Extended and ReadingScreen assessments for any conclusion to be drawn.

The promising findings from this initial trial combined with the RSC's ability to deliver the training at scale has led the Education Endowment Foundation (EEF) to commission an independent

[2] Pupils were given 30 minutes to write a short piece of creative fiction in response to a passage relating to the piece of Shakespeare the intervention group had just completed an RRW session about. Pupils in the control group received the same task without completing the prior RRW session.

efficacy trial of the RRW programme to investigate potential for impact on Year 5 writing attainment. This evaluation is being carried out by the National Foundation for Education Research (NFER).

The trial will comprise an integrated impact evaluation and implementation and process evaluation (IPE), including a longitudinal component. The impact evaluation constitutes a cluster RCT with the allocation of primary schools on a 1:1 basis to intervention and control arms, using school-level randomisation. School-level randomisation is selected to mitigate the risk of contamination while still allowing for adequate power. There will be no stratification for the primary analysis as this is not required to provide valid causal inference. The primary analysis will look to estimate the impact of the programme on short-term pupil writing ability, with secondary analyses investigating impact on vocabulary, grammar, pupil enjoyment of writing, pupil writing self-efficacy and teacher self-efficacy in relation to teaching writing, some of which will also be investigated as possible mediators for the primary outcome. All primary and secondary analyses will be on an intention-to-treat (ITT) basis. In addition, we will assess the effects of dosage and compliance for any impact observed. The trial is designed and powered to detect impact for all pupils (MDES = 0.2) and for pupils eligible for FSM (MDES = 0.219).

The IPE component of the trial will focus on understanding the causal mechanisms at play in the programme to provide context for the findings of the impact analysis. It will also identify lessons relating to programme design and implementation. We will place particular emphasis on the experience of and outcomes for FSM-eligible pupils to explore whether the programme delivers the inclusive learning opportunities it seeks to provide. We will collect and analyse qualitative data on the training provided as well as what the programme looks like in practice via eight in-depth case studies. An endpoint survey will inform us of how common various issues and perspectives are across the trial sample. Usual practice in both intervention and control schools will be established via surveys at baseline and endpoint.

Intervention

For the purpose of this trial, RRW training will be delivered to participating teachers between November 2025 and April 2026, with teachers delivering sessions to their Year 5 pupils following their first training session until June 2026.

A detailed description of the programme in the context of the TIDieR checklist is presented below.

Why: Rationale, theory and/or goal of essential elements of the intervention

RRW aims to improve writing attainment by increasing pupil enjoyment of and motivation to write by using action, gesture and the senses to engage with high-quality texts. The programme seeks to harness a model of ‘embodied cognition’. to channel whole-body creativity into writing. In addition, the programme aims to increase the sophistication of pupils’ vocabulary and syntax through embodied exposure to examples of this in high-quality texts.

Participating teachers receive intensive experiential continuous professional development (CPD) that encourages pedagogical reflection. The programme sessions themselves centre on collaborative and dialogic approaches to exploring and analysing Shakespeare's texts.

Who: Recipients of the intervention

For the purpose of the trial, the direct recipients of the training are qualified teachers of Year 5 classes in state schools in England. Participating teachers must be responsible for delivering at least two English lessons to their Year 5 class per week. Mixed year group classes are also eligible. While in most cases one teacher per school will participate in the trial, a small number of randomly selected schools were permitted a second teacher where requested.

In cases of (mid-year) staff turnover, the new class teacher can join for the remaining training sessions and will be offered additional support by the RSC team. However, the same teacher should participate in all the training sessions wherever possible.

The indirect recipients are the Year 5 pupils in the classes taught by participating teachers.

What: Physical or informational materials used in the intervention

Teachers receive a toolkit of CPD materials, including excerpts from Shakespeare plays and related exercises, in both paper-based and online forms. This toolkit includes example lesson plans for the RRW sessions.

Teachers are also asked to complete a paper-based reflective journal (see following section for more information).

What: Procedures, activities and/or processes used in the intervention

CPD for teachers

The training that teachers receive seeks to replicate the conditions of the classroom, with the teachers playing the part of the pupils to experience the pedagogy from the ground up and understand it from the pupil's perspective. This approach is accompanied by structured moments for pedagogical reflection.

Teachers attend five days of training in total, split into three blocks. Each training block has a different focus:

- Days 1-2: investigating character, exploring scenes and investigating Shakespeare's use of language
- Days 3-4: ways into character and text, exploring dilemmas through text and exploring basic rhetoric
- Day 5: introduction to rhythm, principles of editing and reflections on implementation in the classroom

Each training block starts with 'building the company' activities that seek to create an inclusive, trusting space in which teachers feel at ease. Teachers are encouraged to use similar activities with pupils at the start of each session in the classroom.

A number of different plays are considered across the duration of the training to demonstrate the transferability of different skills and approaches, in place of prescribing a specific curriculum or play text. Teachers are encouraged to reflect on how they would use the various techniques they are learning with other Shakespeare texts as performance and examples of their own choosing. Teachers are also trained to support pupils in using performance as a springboard for writing, particularly through reflective discussion around the pedagogical principles that underpin both the RRW approaches and language development/writing.

The second and third training blocks may occur during programme delivery and opportunities for reflection are offered in relation to implementation and any challenges faced.

Teachers also attend two RSC performances of a Shakespeare play as part of the first two training blocks. This is to help familiarise teachers with Shakespeare's texts and the embodiment of language. Teachers are given tasks to keep in mind when watching the performance, such as reflecting on a particular choice or interpretation that has been made. The performance also serves as a basis for discussions on the following training day.

Finally, pedagogical reflection is supported by use of the reflective journals. Teachers are encouraged to make notes in their journal throughout the training and, once they are back in school, session delivery. The journals provide teachers with prompts to consider and respond to, including what they find surprising and/or challenging in delivering the sessions. Teachers are encouraged to use these notes to feed back to the group at the next training session. In addition, teachers are encouraged to keep a general log of their thinking around developing their practice in relation to the training and tools they have received to support their own professional development. These journals are for teachers' personal use and they are encouraged to take whatever approach to using them that they will find most helpful.

RRW sessions in the classroom

The RRW delivery is unscripted, in that the training aims to provide teachers with the skills and knowledge that they can then apply as they decide in their own classrooms. The training gives teachers experience of activities and techniques that combine collaborative, dialogic and oracy-based learning with role-playing, close reading and textual interpretation. The teachers are then encouraged to use these activities and techniques in sessions with their Year 5 pupils.

When delivering the sessions, teachers are encouraged to use group work to support pupils in interpreting the text.

Each session in the classroom will typically involve a warm-up exercise and introduction of the chosen Shakespeare text, followed by between two and four RRW exercises over the course of approximately one hour. Many of these exercises would include small-group and pair work.

Examples of RRW exercises include:

- Choral and whispered readings
- Creation of soundscapes
- Creating tableaux
- Improvising in character
- Scaffolded analysis of language that starts with a word, then a line, then a whole speech
- Discussing ideas with peers

Several performance exercises lead directly into writing exercises.

While it is possible for teachers to use RRW techniques in other subjects and with a range of (non-Shakespeare) texts, they are designed for and work best with Shakespeare texts. As a result, only sessions focused on Shakespeare will be considered for the purpose of the trial.

Who: Intervention providers / implementers

The CPD is delivered by four RSC CPD Associate Learning Practitioners (ALPs) who already deliver CPD for using rehearsal room techniques to teachers on behalf of the RSC. All CPD ALPs have at least two years of experience as an RSC ALP, working directly with young people. Many ALPs also have experience working in schools and/or with teachers beyond the RSC. In addition, CPD ALPs complete a structured RSC training programme comprised of observations and working with a mentor. The CPD ALPs will not receive any additional training for the purpose of the trial but will be undertaking planning sessions to adapt their usual CPD for Year 5 teachers and to address writing outcomes where feasible.

The CPD ALPs will deliver the training sessions in pairs with a maximum of 25 teacher participants per session.

How: Mode of delivery

Both the CPD and the programme sessions are delivered in-person.

Where: Location of the intervention

The training days take place in the RSC rehearsal rooms in Stratford-upon-Avon. For this trial, teachers will receive free accommodation while attending the training days and will be compensated for their travel to attend the training.

Programme sessions are designed to be delivered in the classroom.

When and how: Duration and dosage of the intervention

Teachers participate in five full days (up to 6 hours per day) of in-person training, in blocks of two-two-one. The first two days take place in the autumn term, after which teachers are permitted to

start delivering the programme sessions in their schools. Training days three and four take place in the spring term, and the final day in the summer term. Each of the first two training blocks include attendance at an RSC performance of a Shakespeare play in the evening (up to 3 hours).

Ideally teachers will attend all five days but attending a minimum of three days, including the first two training days, would still be deemed acceptable for programme delivery.

For the purpose of the trial, each teacher should deliver a minimum of 20 hours of lesson time to their Year 5 class using RRW approaches between November 2025 and June 2026. Teachers are advised to embed these sessions within their English curriculum. No prescriptions are made regarding when these hours occur to enable teachers to find the best fit within their curriculum and the structures of their school.

As the sessions are expected to take place during usual class time there are no additional specifications around pupil attendance.

Tailoring: Adaptation of the intervention

The RRW programme has been designed to be suitable for mixed ability groups, including for pupils with special educational needs and/or disabilities (SEND) and/or English as an additional language (EAL). Opportunities to reflect on and discuss how delivery of the programme could be tailored to suit different pupils' needs and school settings are provided as part of the training.

This is not a scripted intervention, and teachers are actively encouraged to develop their own RRW sessions, not just replicate the ones they did as part of the CPD. The aim is that teachers feel they have ownership over these sessions, adapting them to meet their own context and pupil needs. Timing of the sessions is likewise flexible, with the only stipulation being a minimum of 20 hours delivered over the trial period.

How well (planned): Strategies to maximise effective implementation

The RSC has sought to maximise the effectiveness of programme implementation by providing training that is:

(a) Delivered by expert practitioners who have experience of rehearsal room practices and of delivering high-quality CPD; and

(b) Staggered across the academic year to consolidate learning alongside delivery.

In addition, the RSC provides a number of detailed supporting materials, including the Primary Toolkit, online resources, and lesson plans focused on writing activities for the pupils.

How well (planned): Evidence of implementation variability

While the programme encourages teachers to adapt the pedagogy to suit their professional identity, their student groups, and the curricula in which these techniques are embedded, it is expected that a drama-based approach to teaching Shakespeare is adopted. As such, reading aloud from Shakespeare's text is not sufficient. In order for pupils to fully experience and embody

Shakespeare's language and characters, they have to be active, collaborative, and away from their desks. Other variations in implementation are, however, more generative. Historically, these have included the application of drama techniques to non-Shakespearean texts in the English classroom, alongside variations in delivery scheduling. Some schools implement this work intensively, for example during a 'Shakespeare Fortnight' or as part of a performance output. Others embed RRW throughout their English curriculum and often adapt it for specific areas of literacy.

Theory of Change

The Theory of Change (ToC) for the RRW programme is shown in Figure 1. It outlines the target population (Year 5 teachers and their pupils), and the activities, outputs and short-term outcomes intended to result in improved creative writing performance among participating pupils. An additional strand of outputs and outcomes reflects the intended trajectory for teacher learning, outcomes and broader impact.

The RRW sessions offer learning opportunities along three causal pathways. The first is centred on pupil oracy and performance leading to outcomes such as enjoyment and motivation in relation both to the sessions themselves and writing specifically. The second pathway focuses on the experience of the text itself, exploring and interpreting its meaning and what it is trying to achieve. The outcomes stemming from this concern the use of language and awareness of the choices involved in composing a text. These two pathways combine the experience of performance with exploration of a text to feed into a third pathway that focuses on embodied exposure to new vocabulary and syntactical structures that subsequently make their way into the pupils' work. These three pathways all feed into the primary outcome of improved quality of pupils' creative writing, leading into the longer-term objective of improved pupil writing across the curriculum.

The short-term outcomes resulting from each of these pathways – enjoyment, motivation, vocabulary and syntactical development – all represent outcomes of interest in their own right. However, they are also theorised to be mediators for the primary outcome of writing attainment. This evaluation will test these secondary outcomes and, for pupil enjoyment and self-efficacy, also their relationship to the primary outcome. This will make it possible to identify which of these two pathways plays the most significant role in any impact in relation to the primary outcome that may be observed. The results of this analysis will enable further development of the ToC to present a clearer picture of the mechanisms at play.

Given the oracy-based nature of many of the RRW activities, we have theorised that this will support pupils to grow their confidence around using spoken language. This will be explored as a standalone secondary outcome as part of the IPE using a measure of feelings about communicating with other people (see the IPE design section for more details). Confidence in relation to spoken language is one of the key objectives for language and literacy in the Primary National Curriculum (Department for Education, 2013).

In parallel to the pupil pathways there runs a pathway of activities, outputs and outcomes concerning the teachers. As RRW is primarily a CPD programme, it is expected that the teachers'

experience will play an important role in shaping that of their pupils. The activities and outputs outline not just the training activities but also the role of the pedagogical discussion and reflection that occurs around this. The unscripted nature of the programme sessions encourages teacher autonomy, which in turn requires quite a high level of engagement and ability on the teacher's part. As a result, the short-term teacher outcomes of confidence, enjoyment and motivation in relation to engaging with Shakespeare's texts and using rehearsal room techniques in their practice are understood to be a pre-condition for successful delivery of the sessions that constitute the activities for the pupil participants.

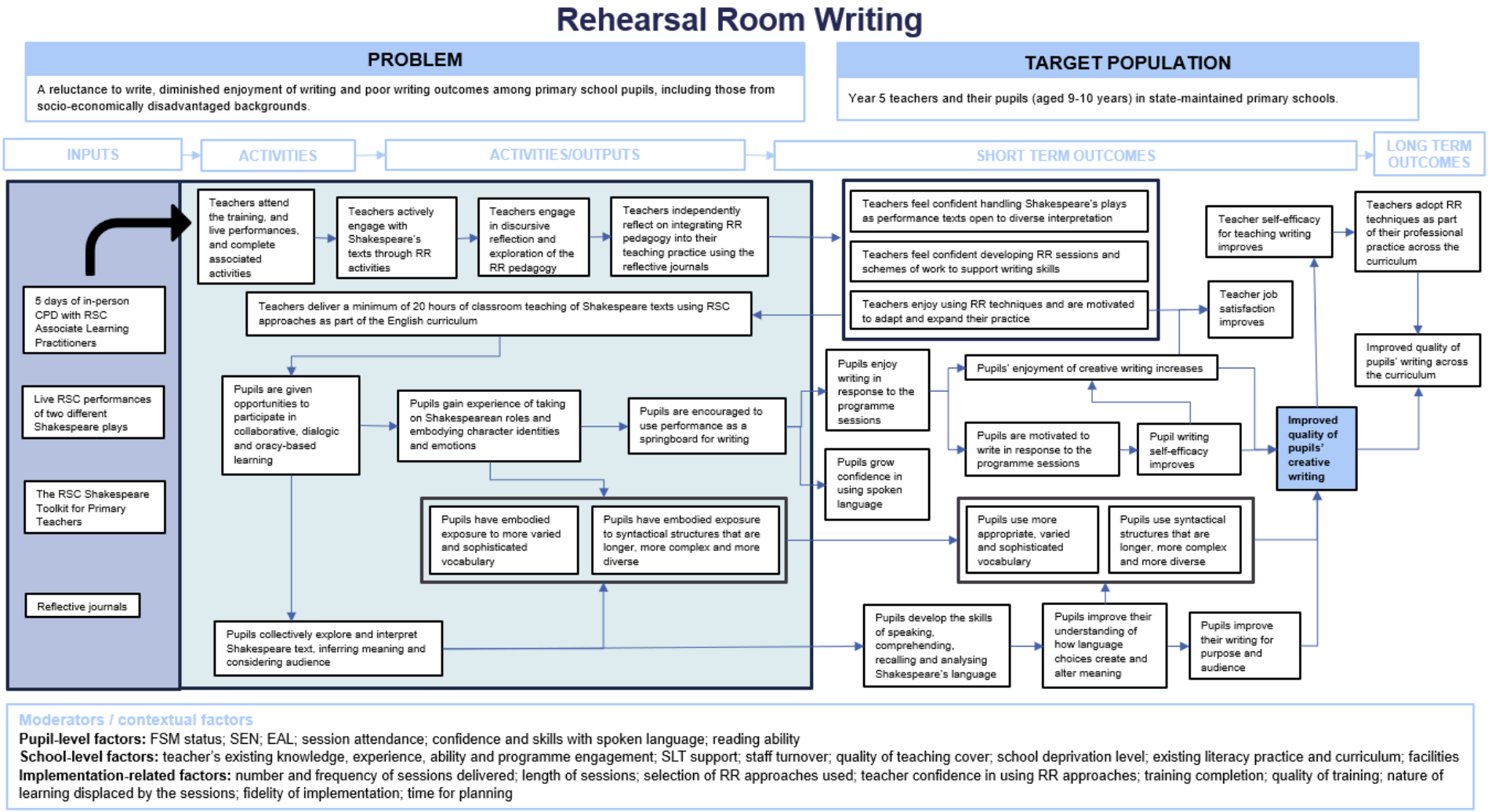
It is theorised that, should the rehearsal room techniques be effective for improving pupil writing, teacher self-efficacy in relation to teaching writing will increase, as will their motivation to adopt those techniques beyond the trial,[3] creating a virtuous cycle of implementation. It is also possible that teachers may feel motivated to adopt these techniques for use in other areas of the curriculum, which would (if effective) increase the likelihood of pupil writing across the curriculum improving as well, but it is not within the scope of this trial to measure this.

There is evidence to suggest that interpersonal relations at the classroom level are a key driver of teacher job satisfaction (Lopes and Oliveira, 2020; Harrison, King and Wang, 2023). This has led us to theorise that the mutual enjoyment teachers and pupils are intended to draw from the RRW sessions may result in higher job satisfaction for the teachers. The association between professional development and job satisfaction for teachers (Smet, 2022) further supports this hypothesis.

Finally, a number of potential moderators and influential contextual factors have been identified for consideration of their effect, in relation to both pupil and school characteristics and the nature of programme implementation. Given that the programme does not focus on building foundational literacy, it is likely that confidence and skills with spoken language and reading ability may affect the extent to which a pupil can draw benefit from the RR activities. The unscripted and relatively unstructured nature of the programme means that teacher experience, ability and programme engagement will likely moderate the impact the programme can achieve for their pupils primarily via quality of delivery. The extent to which the RRW techniques differ from the teacher's usual practice will also influence the magnitude of any pupil impact observed. Programme dosage in terms of number, frequency and length of sessions will be expected to play a moderating role. School context, including any staffing challenges and levels of support from the senior leadership team (SLT) for implementing new techniques may also impact implementation.

[3] Evidence indicates a close if not yet fully determined relationship between the constructs of self-efficacy and motivation (depending on how they are measured). See for example Williams, Kessler and Williams, 2014; Williams and Rhodes, 2016; Schunk and DiBenedetto, 2021.

Figure 1: Theory of Change



Impact evaluation design

Research questions

Primary

RQ1: How effective is RRW at improving the writing ability of Year 5 pupils?

Secondary

RQ2a: How effective is RRW at increasing the richness of Year 5 pupils' written vocabulary?

RQ2b: How effective is RRW at increasing the sophistication of Year 5 pupils' written vocabulary?

RQ3a: How effective is RRW at increasing the complexity of Year 5 pupils' use of noun phrases in writing?

RQ3b: How effective is RRW at increasing Year 5 pupils' use of relative clauses to complexify noun phrases in writing?

RQ4: How effective is RRW at improving the writing ability of Year 5 pupils eligible for FSM?

RQ5: How effective is RRW at improving Year 5 pupils' enjoyment of writing?

RQ6: How effective is RRW at increasing Year 5 pupils' writing self-efficacy?

RQ7: How effective is RRW at improving the self-efficacy of Year 5 teachers in relation to teaching writing?

RQ8a: To what extent is pupil enjoyment of writing a mediator of the primary outcome?

RQ8b: To what extent is pupil writing self-efficacy a mediator of the primary outcome?

RQ9: How does the impact of RRW on pupils' writing attainment vary by the number of sessions taught by the teacher?

Longitudinal follow-up RQs:

RQ10a: What is the impact of RRW on pupils' writing attainment one year after the end of programme delivery (end of Year 6)?

RQ10b: What is the impact of RRW on the writing attainment of pupils eligible for FSM one year after the end of programme delivery (end of Y6)?

RQ11a: What is the impact of RRW on pupils' reading attainment one year after the end of programme delivery (end of Year 6)?

RQ11b: What is the impact of RRW on the reading attainment of pupils eligible for FSM one year after the end of programme delivery (end of Year 6)?

RQ12: In RRW schools, is pupil performance in maths at the end of Year 6 (one year after the end of the programme) no worse than for pupils in the control group? (non-inferiority analysis)

These RQs are explained briefly below:

The primary research question (**RQ1**) focusses on holistic writing ability, assessed through comparative judgement.

RQ2 focuses on the impact of the intervention on the maturity of children's written language in terms of vocabulary use. Mature vocabulary use will be evaluated in terms of two measures which previous research has suggested correlate with children's age and teachers' evaluations of writing quality (Durrant, Brenchley and McCallum, 2021), specifically, the diversity of vocabulary use (**RQ2a**) and the percentage of words that are not high-frequency in English (**RQ2b**).

In contrast, **RQ3** focuses on the impact of the intervention on the maturity of children's written language in terms of syntax. It will focus on development within the noun phrase, which previous research suggests is a key site of development in child writing (Durrant and Brenchley, 2023). **RQ3a** looks at the overall complexity of noun phrases, as measured by the mean number of words per noun phrase. **RQ3b** looks at relative clauses as a specific element within noun phrase complexity, as previous research has suggested that an increasing number of relative clauses is a key marker of maturity (Durrant and Brenchley, 2023).

RQ4 is a subgroup analysis of the primary research question (**RQ1**). The subgroup of interest is pupils who have been recorded as eligible for FSM in the last 6 years.

RQ5 and **RQ6** focus on the effect of the intervention on pupils' enjoyment of writing and pupils' writing self-efficacy, assessed through a pupil survey.

RQ7 focuses on the effect of the intervention on teachers' self-efficacy in relation to teaching writing, assessed through a teacher survey.

The extent to which the pupil survey outcomes analysed in RQ5 and RQ6 mediate the primary outcome variable is explored in **RQ8**.

RQ9 is a dosage analysis that examines the relationship between the primary outcome and the number of sessions delivered.

One year after the end of the intervention, a follow-up analysis will investigate any long-term differences between the intervention and control groups using KS2 outcomes in writing (**RQ10**), reading (**RQ11**) and maths (**RQ12**). The FSM subgroup will also be assessed for writing and reading. While reading is not a focus of the main trial, qualitative data from the earlier RCT (McCulloch and Collins, 2023) indicates the possibility of improved reading comprehension and inferencing ability that we felt merited further attention. We are also looking at the effect on maths outcomes due to previous evidence of writing interventions leading to negative impacts on other subject areas, likely due to learning displacement effects (Torgerson *et al.*, 2018). The

hypothesis for the possible negative impact of the intervention on maths makes a non-inferiority approach appropriate.

Design

Table 1: Trial design

Trial design, including number of arms		Two-arm, school randomised
Unit of randomisation		School
Stratification variables (if applicable)		None
Primary outcome	**Variable**	Writing Ability
	Measure (instrument, scale, source)	Writing Assessment Measure – Comparative Judgement
Secondary outcome(s)	**Variable(s)**	i) Lexical diversity ii) Lexical sophistication iii) Noun phrase complexity iv) Pupils' enjoyment of writing v) Pupils' writing self-efficacy vi) Teacher self-efficacy in relation to teaching writing vii) KS2 writing viii) KS2 reading ix) KS2 maths
	Measure(s) (instrument, scale, source)	i) Moving average type token ratio ii) Percentage of lemmas falling outside the most frequent 2,000 and the most frequent 3,000 words in English iii) Mean number of words per noun phrase and frequency of relative clauses iv) Liking Writing Scale v) Self-efficacy for Writing Scale vi) Teacher Efficacy Scale for Writing vii) KS2 writing viii) KS2 reading ix) KS2 maths
Baseline for primary outcome	**Variable**	Writing Ability
	Measure (instrument, scale, source)	Writing Assessment Measure – Comparative Judgement
Baseline for secondary outcome	**Variable**	i) Lexical diversity ii) Lexical sophistication iii) Noun phrase complexity iv) Pupils' enjoyment of writing v) Pupils' writing self-efficacy vi) Teacher self-efficacy in relation to teaching writing vii) Writing ability (same as BL for primary outcome), viii) KS1 age related expectations in reading

		ix)	KS1 age related expectations in maths
	Measure (instrument, scale, source)	i)	Moving average type token ratio
		ii)	Percentage of lemmas falling outside the most frequent 2,000 and the most frequent 3,000 words in English
		iii)	Mean number of words per noun phrase and frequency of relative clauses
		iv)	Liking Writing Scale
		v)	Self-efficacy for Writing Scale
		vi)	Teacher Efficacy Scale for Writing
		vii)	Writing Assessment Measure – Comparative Judgement
		viii)	KS1 age related expectations in reading
		ix)	KS1 age related expectations in maths

This is an efficacy trial to evaluate the impact of the Rehearsal Room Writing programme. The study is a two-arm RCT randomised at school level. Outcomes will be measured in Year 5 pupils taught by participating teachers. Teachers in schools which are randomised to the intervention arm will deliver the RRW programme as described in the Intervention section of this document. Teachers in schools which are randomised to the control arm will deliver business as usual. The primary outcome in this evaluation will be the Writing Assessment Measure assessed using comparative judgement (WAM_CJ). Secondary outcomes include the richness and sophistication of written vocabulary, complexity of noun phrases in writing, use of relative clauses to complexify noun phrases in writing, pupils' enjoyment of writing, pupils' writing self-efficacy, self-efficacy of teachers in relation to teaching writing, KS2 writing teacher assessment, KS2 scaled score in reading and KS2 scaled score in maths.

Participant selection

Participants in this study will be qualified teachers working in state maintained primary schools in England who plan to teach at least two English lessons per week to a Year 5 class, and the pupils in the participating teacher's Year 5 class. As part of the recruitment process, we will check that participating teachers have not previously undertaken any RSC training; any who have will be ineligible to participate in the trial. Special schools will not be eligible to participate as they serve a specific cohort of pupils that the programme was not designed for. Schools that are participating in The Write Journey Evidence into Action Partnership or any of the following EEF 2025/26 writing trials in will also be ineligible: Pathways Literacy, Power of Reading, and Writing Roots. This is due to a concern around the burden that participation in multiple trials will place on a school and conflicting priorities, as well as the risk of confounding the results.

All pupils within the participating teacher's Year 5 class will be eligible and will be included in the primary analysis. Some secondary outcome measures (described below) require scripts to be transcribed (RQ2) or are time intensive to complete (RQ3). To ensure reasonable allocation of budget among research questions, subsets of pupils will be randomly selected within each school to be analysed for these research questions.

Outcome measures

Primary outcome

The primary outcome measure will be a Writing Assessment Measure assessed using comparative judgement (WAM_CJ). We will work with No More Marking[4] (NMM) to generate this measure. The measure aligns with the long-term outcome theorised by the logic model. Assessment of writing is multi-dimensional and contested (Clarkson, 2024), and the most commonly used form of assessment, rubric-based scoring, has been shown to be less accurate and consistent than the increasingly used Comparative Judgement (Pollitt, 2012; Pinot de Moria, Wheadon and Christodoulou, 2022). Pupils will be presented with a narrative writing pictorial prompt at both baseline and endpoint (a picture from the NMM archive, unrelated to Shakespeare). They will be asked to write for up to 30 minutes.

Completed baseline and endpoint assessments will be scanned and loaded into the NMM platform. A team of judges, recruited and managed by NFER, will be trained to evaluate the scripts using a 'comparison' approach. Judges will be presented with two assessment scripts and asked to judge which is the better piece of writing. Judges will be allocated a quota of comparisons to make, with the NMM platform ensuring that each script is judged/compared with a minimum of 10 other scripts, across multiple judges.

To enable monitoring of judges performance, NMM will provide NFER with data on (i) each judge's median time to make a judgment, (ii) % left click and (iii) judge infit. This data will be monitored during the judging period and if necessary used to inform additional training requirements or changes to the allocation of judges during the marking window. More information about how these metrics will be monitored follows:

- **Median Time** - the median time taken by a judge in their judgements on this task. Quite short median times might indicate that a judge is rushing through their judgements (it is informative to compare the median time to the infit score), and a long median time might indicate that the judge is facing some problems with the judgement process.

- **% Left Click** - In the judging screen, scripts are presented randomly on the left or right hand side of the screen. Therefore, we would generally expect that the percentage of times that a judge clicks in favour of the 'Left' script would be around 50%. If the % Left Click value is much higher or lower than 50%, it might indicate that the judge is not thinking carefully about their judgements and just staying with the left or right hand side. Again, it is informative to compare this value with the infit statistic and the median time taken.

- **Infit** - This statistic is an indication of the degree of consistency or inconsistency of this judge's decisions compared to those of the other judges. A low infit score indicates better consistency, a higher infit score indicates worse consistency.

[4] https://www.nomoremarking.com/?countryCode=GB

Once the marking activity is complete, NMM will provide NFER with pupils' scaled scores which will be converted to a writing age for each pupil. NMM will also provide a task reliability figure (between 0-1) which is an indicator of the reliability/trust in final scores.

Secondary outcomes

Richness of vocabulary use refers to two constructs: lexical diversity and lexical sophistication. Lexical diversity (**RQ2a**) refers to the extent to which writers used a range of different words, rather than repeating items. This is known to correlate with both writer age and teachers' evaluation of writing quality (Durrant, Brenchley and McCallum, 2021). Lexical diversity will be measured using the moving average type token ratio. Research has shown that this provides a robust way of quantifying the number of different words that writers used independently of text length (Zenker and Kyle, 2021), so enabling fair comparison of different texts.

Lexical sophistication (**RQ2b**) refers to the extent to which writers use vocabulary that is not frequent in the language as a whole. This also correlates with both writer age and teachers' evaluation of writing quality (Durrant, Brenchley and McCallum, 2021). Lexical sophistication is often operationalized in terms of the percentage of words that do not appear on a list of high-frequency vocabulary. For greater sensitivity, two such measures will be used in this project: the percentage of lemmas falling outside the most frequent 2,000 and the most frequent 3,000 words in English, as evidenced by the most recent edition of the British National Corpus (Brezina, Hawtin and McEnery, 2021).

Noun phrase complexity (**RQ3**) refers to the average number of components included within each noun phrase. This is operationalized as the mean number of words per noun phrase (**RQ3a**), a measure which has been shown to correlate with learner age (Durrant and Brenchley, 2023). Previous research has highlighted the use of relative clauses as a key site of complexity development within noun phrases (Durrant and Brenchley, 2023), so the frequency of this form is studied separately as a potential area of development (**RQ3b**). Noun phrases and their internal components will be initially identified by an automated parser (Manning *et al.*, 2014). The output of the parser will then be manually checked for accuracy by analysts who will be trained by the project team.

RQ4, **RQ8a**, **RQ8b** and **RQ9** will use the primary outcome variable of WAM_CJ as described above.

The pupil surveys at baseline and endpoint will include items from the Self-Efficacy for Writing Scale and items from the Liking Writing Scale (Bruning et al., 2013). These are existing validated scales, suitable for this age group. Questions will be read aloud to pupils by the teacher. Our initial plan, to apply the factor loadings in Bruning et al., 2013, was revised due to small wording changes to the questions which were considered essential for appropriate use in a UK classroom. Under the present plan, baseline responses to the items on each of these scales will undergo (separate) factor analyses and the factor from each which explains the most variance will be the baseline outcome for RQ5 and RQ6. The loadings from these factor analyses will be applied to the endpoint survey responses to form the outcome measures for **RQ5** (pupils' enjoyment of writing) and **RQ6** (pupils' writing self-efficacy).

The outcome measure for **RQ7** is teacher self-efficacy in relation to teaching writing. In the teacher surveys at baseline and at endpoint, teachers will be asked four items from the teacher efficacy for improving students' writing performance scale (Gilbert and Graham, 2010). Responses to the items on this scale at baseline will undergo a factor analysis and the factor which explains the most variance will be the baseline outcome for RQ7. The loadings from this factor analysis will be applied to the endpoint responses to form the outcome measures for RQ7.

The outcomes for **RQ10a**, **RQ10b**, **RQ11a**, **RQ11b** and **RQ12** are all KS2 outcomes held in the National Pupil Database. For **RQ10a** and **RQ10b**, the outcome for the writing teacher assessment (KS2_WRITTAOUTCOME) will be used to create a binary outcome measure where 1 indicates that the pupil is either 'working at the expected standard' or 'working at greater depth within the expected standard' (i.e. 1 indicates KS2_WRITTAOUTCOME is 'EXS' or 'GDS', 0 indicates any other code recorded for this variable).. For **RQ11a** and **RQ11b**, the scaled score in reading (KS2_READSCORE) will be used as the outcome measure. For **RQ12**, the scaled score in maths (KS2_MATSCORE) will be used as the outcome measure.

Baseline measures

All endpoint measures for RQ1 to RQ9 will also be recorded at baseline before randomisation via a baseline writing assessment (also from NMM; this will use a different pictorial prompt to the one used at endpoint) and baseline pupil and teacher surveys. Scores will be used as covariates in the appropriate analysis models. Baseline scores will not be used for stratification at randomisation (more detail in Randomisation section). For RQ10a, RQ10b, RQ11a and RQ11b we will use the WAM_CJ baseline score from NMM, as in RQ1. For RQ12, Year 4 Multiplication Tables Check score (MTC_FormMark) will be used as a baseline measure for each pupil.

Sample size

Table 2: Sample size calculations – analysis numbers after anticipated attrition

		Primary (RQ1)	FSM (RQ4)	Secondary (RQ2)
Measure		WAM_CJ	WAM_CJ	MATTR50
Minimum Detectable Effect Size (MDES)		0.200	0.219	0.178
Pre-test/ post-test correlations	level 2 (school)	0.68	0.68	0.5
Intracluster correlations (ICCs)	level 2 (school)	0.19	0.19	0.09
Alpha		0.05	0.05	0.05
Power		0.8	0.8	0.8
One-sided or two-sided?		Two-sided	Two-sided	Two-sided
Average cluster size		26.0	7.6	9.3
Number of schools	Intervention	82	82	82
	Control	82	82	82

	Total	164	164	164
Number of pupils	Intervention	2133	621	762
	Control	2133	621	762
	Total	4266	1242	1524

Sample size calculations were undertaken using the *PowerUpR* package in R statistical software. For the primary measure, the ICC and pre-post correlation are taken from the Year 5 cohort in the Helping Handwriting Shine trial which used a similar measure (Stone *et al.*, 2022). The average number of pupils per school is the average KS2 class size in 23/24 (28) taken from data published by the Department for Education (DfE) (Department for Education, 2024b) multiplied by the anticipated number of teachers per school (1.09) and adjusted for expected pupil attrition (15%).

For the FSM subset calculation, the primary analysis ICC and pre-post correlation are used. The average number of pupils per class is 29.1% of the number of pupils expected in the primary measure. This is the FSM percentage among all Y5 pupils in 23/24, taken from published DfE data (Department for Education, 2024b).

Since only a subset of pupils will be analysed for some of our secondary measures, we have also undertaken a power calculation for one secondary measure (moving average type token ratio) to determine the size of the subset. The ICC was determined through analysis of similar data, supplied by the University of Exeter. Unfortunately, no data was available for pre-post correlation for this calculation. Existing literature demonstrates that correlations of around 0.7 are common in educational research (Singh *et al.*, 2023). We have selected 0.5 for this calculation as a conservative but not unrealistically small estimate.

The number of schools and pupils shown in the table are the numbers expected to be analysed. We anticipate 15% pupil attrition and 10% school attrition so to achieve the MDES above we aimed to recruit **200** teachers from **183** schools (**5,600** pupils). Near the end of recruitment (June 2025) it was noted that the average number of pupils per teacher in the recruited schools was lower than anticipated (26.6 vs. 28). To accommodate this difference while still maintaining the total number of teachers (for MDES of 0.2, power 80%), we updated the recruitment target to be **200** teachers from **184** schools (**5,320** pupils), i.e. 16 schools with two teachers in the trial instead of 17.

Randomisation

The randomisation will be school level randomisation with equal allocation between groups. No stratification is implemented so simple randomisation will be applied. Randomisation will occur after recruitment has closed and after teachers have been nominated by schools.

The subset of pupils that will be analysed for RQ2 and RQ3 will be randomly selected from the pupils within each class. Sampling will be stratified by FSM eligibility i.e. it will be performed separately within FSM eligible pupils and pupils not known to be eligible for FSM, with a proportionate number of pupils selected in each category to ensure that the percentage of FSM pupils is accurately represented in the subset of pupils. Due to timeline, the FSM variable used

for this stratification will be collected from schools and will be current FSM eligibility, in contrast to the FSM variable used for the subgroup analysis (RQ4) which will be sourced from the NPD and will be 'recorded as eligible for FSM in the last 6 years'.

Randomisation will be done using R statistical software and a seed will be set for reproducibility. Code will be included in the appendix of the final report.

Statistical analysis

The main analyses will be intention-to-treat and will follow the EEF Statistical Analysis Guidance (EEF, 2022b).

Analysis for RQ4, RQ10a, RQ10b, RQ11a, RQ11b and RQ12 require data from the NPD (FSM eligibility, KS2 outcomes). These analyses will be performed in the ONS Secure Research Service (SRS) or, if available, in the Integrated Data Service (IDS). All other analyses will be performed locally.

Primary analysis

The primary outcome measure of WAM_CJ will be used as the dependent variable in a linear mixed effects model with intervention group as a predictor, controlling for baseline scores and class as covariates, and accounting for clustering of pupils within schools. The analysis population will be all pupils included in the evaluation. The comparison between the intervention and control groups will answer RQ1.

Secondary analysis

RQ2a: The analysis population will be a randomly selected subset of 10 pupils per class. The secondary outcome measure of moving average type token ratio will be used as the dependent variable in a linear mixed effects model with intervention group as a predictor, controlling for baseline scores, FSM (stratification variable used in selection of the sub-sample for this analysis) and class as covariates, and accounting for clustering of pupils within schools. The comparison between the intervention and control groups will answer RQ2a.

RQ2b: The analysis population will be a randomly selected subset of 10 pupils per class (the same subset as RQ2a). The secondary outcome measures of the percentage of lemmas falling outside the most frequent 2,000 and the most frequent 3,000 words in English will be used as dependent variables in two linear mixed effects model with intervention group as a predictor, controlling for baseline scores, (stratification variable used in selection of the sub-sample for this analysis) and class as covariates, and accounting for clustering of pupils within schools. The comparison between the intervention and control groups in the two models will answer RQ2b.

RQ3a: The secondary outcome measure of mean number of words per noun phrase will be used as the dependent variable in a linear model with intervention group as a predictor, controlling for baseline scores. The analysis population will be one randomly selected pupil per class. The comparison between the intervention and control groups will answer RQ3a.

RQ3b: The secondary outcome measure of frequency of relative clauses will be used as the dependent variable in a linear model with intervention group as a predictor, controlling for baseline scores. The analysis population will be one randomly selected pupil per class. The comparison between the intervention and control groups will answer RQ3b.

RQ4: The primary outcome of WAM_CJ will be used as the dependent variable in a linear mixed effects model with intervention group as a predictor, controlling for baseline scores and class as covariates, and accounting for clustering of pupils within schools. The analysis population will be pupils who have been recorded as eligible for FSM in the last 6 years. An additional model will be run repeating RQ1 but with the addition of an interaction term of intervention group by FSM eligibility.

RQ5: The secondary outcome measure of pupils' enjoyment of writing will be used as the dependent variable in a linear mixed effects model with intervention group as a predictor, controlling for baseline scores and class as covariates, and accounting for clustering of pupils within schools. The analysis population will be all pupils included in the evaluation. The comparison between the intervention and control groups will answer RQ5.

RQ6: The secondary outcome measure of pupils' writing self-efficacy will be used as the dependent variable in a linear mixed effects model with intervention group as a predictor, controlling for baseline scores and class as covariates, and accounting for clustering of pupils within schools. The analysis population will be all pupils included in the evaluation. The comparison between the intervention and control groups will answer RQ6.

RQ7: The secondary outcome measure of teacher self-efficacy in relation to teaching writing (average value per school) will be used as the dependent variable in a linear model with intervention group as a predictor, controlling for baseline scores. The analysis population will be all schools included in the evaluation. The comparison between the intervention and control groups will answer RQ7.

RQ8a: The secondary outcome of pupils' enjoyment of writing will be explored as the mediator in a basic mediation model (Pieters, 2017), with the primary outcome of WAM_CJ as the dependent variable. The analysis population will be all pupils included in the evaluation. The percentage of the total effect made up by the indirect effect will be reported to answer RQ8a.

RQ8b: The secondary outcome of pupils' writing self-efficacy will be explored as the mediator in a basic mediation model, with the primary outcome of WAM_CJ as the dependent variable. The analysis population will be all pupils included in the evaluation. The percentage of the total effect made up by the indirect effect will be reported to answer RQ8b.

RQ9: The primary outcome of WAM_CJ will be used as the dependent variable in a linear mixed effects model with number of sessions as a predictor, controlling for baseline scores and class as covariates, and accounting for clustering of pupils within schools. The analysis population will be all intervention group pupils included in the evaluation. The coefficient of the dosage variable will be reported to answer RQ9.

RQ10a: The secondary outcome measure of KS2 writing outcome (binary measure where 1 indicates that the pupil is either 'working at the expected standard' or 'working at greater depth within the expected standard') will be used as the dependent variable in a binomial generalised linear mixed effects model with intervention group as a predictor, controlling for baseline scores and class as covariates, and accounting for clustering of pupils within schools. The analysis population will be all pupils included in the evaluation. The comparison between the intervention and control groups will answer RQ10a.

RQ10b: The secondary outcome measure of KS2 writing outcome (binary measure where 1 indicates that the pupil is either 'working at the expected standard' or 'working at greater depth within the expected standard') will be used as the dependent variable in a linear mixed effects model with intervention group as a predictor, controlling for baseline scores and class as covariates, and accounting for clustering of pupils within schools. The analysis population will be pupils who have been recorded as eligible for FSM in the last 6 years. The comparison between the intervention and control groups will answer RQ10b.

RQ11a: The secondary outcome measure of KS2 reading scaled score will be used as the dependent variable in a linear mixed effects model with intervention group as a predictor, controlling for baseline scores and class as covariates, and accounting for clustering of pupils within schools. The analysis population will be all pupils included in the evaluation. The comparison between the intervention and control groups will answer RQ11a.

RQ11b: The secondary outcome measure of KS2 reading scaled score will be used as the dependent variable in a linear mixed effects model with intervention group as a predictor, controlling for baseline scores and class as covariates, and accounting for clustering of pupils within schools. The analysis population will be pupils who have been recorded as eligible for FSM in the last 6 years. The comparison between the intervention and control groups will answer RQ11b.

RQ12: The secondary outcome measure of KS2 maths scaled score will be used as the dependent variable in a linear mixed effects model with intervention group as a predictor, controlling for baseline scores and class as covariates, and accounting for clustering of pupils within schools. The analysis population will be all pupils included in the evaluation. The comparison between the intervention and control groups, with a 'not inferior to' hypothesis, will answer RQ12.

Estimation of effect sizes

Effect size will be calculated as Hedge's g effect sizes, using total variance from a model without covariates. 95% confidence intervals will be computed for all reported effects.

Analysis in the presence of non-compliance

We will use three definitions of compliance to create three school-level compliance metrics. Minimal compliance in training will be a binary measure defined as having attended at least three days of training including the first two days. Optimal compliance in training will be a binary

measure defined as having attended all five training days. Where more than one teacher was nominated by the school, both teachers must achieve the definition for the school to be considered compliant. Compliance in delivery will be a continuous measure defined as the number of hours delivered to pupils. For each of the compliance metrics, a complier average causal effect (CACE) analysis will be undertaken using a two stage least squares instrumental variable approach. These three analyses will be independent of each other. We recognise that the optimal school training compliance measure risks violating the CACE analysis exclusion restriction as we would expect at least some effect of teachers attending fewer than 5 training sessions. The appropriate caveats will be included in reporting and the primary focus for the teacher training compliance will be the minimal compliance measure.

Missing data analysis

The number and proportion of pupils with missing endpoint WAM_CJ scores (primary outcome variable) will be reported. If the percentage of pupils (present at randomisation) missing this outcome variable is less than 5%, no further missing data analysis will take place. If the percentage of pupils missing is greater than 5%, a logistic multilevel model will be run with a binary outcome variable indicating missing endpoint WAM_CJ. This model will include the intervention group variable as a predictor, along with pupil, teacher and school characteristics (to be defined in the statistical analysis plan [SAP]). Any of the additional variables which demonstrate a significant association with missingness will be included as a covariate in a rerun of the RQ1 analysis as a sensitivity check.

Implementation and process evaluation (IPE) design

This IPE approach was confirmed as part of the set-up phase following the IDEA workshop, which provided the evaluation team with the opportunity to discuss the nature and theory of the RRW programme with the RSC team in-depth. This informed the evaluation team's understanding of the key elements of interest for the IPE, which has been designed to support understanding of the impact findings as well as how the programme could potentially be improved in future.

A key focus of this evaluation will be to better understand the causal mechanisms at play in the programme, including via secondary outcomes, to provide context for the results from the impact analysis. The limited extent to which the previous RSC trial explored the challenges of programme implementation means that this will be a particular focus for this IPE. We will also explore any unintended consequences, particularly additional burden for teachers (including attendance at the five training days) and displacement of learning. In addition to teacher and pupil perceptions about the programme, we will capture the extent to which the techniques taught in the training are integrated into the teachers' practice, and their intentions to sustain this in the longer term.

Finally, we will explore the experience of and outcomes for FSM-eligible pupils to understand whether the programme delivers the inclusive learning opportunities it seeks to provide.

Research questions

Fidelity, adaptation & reach

> **IPE RQ1:** *What did implementation of RRW look like in practice?*

- Did teachers attend the training?
- How many RRW sessions did teachers deliver, at what intervals and how long did the sessions usually last? What other lesson(s) did the RRW sessions replace?
- What did the RRW training look like in practice**?** How well was it seen to enable teachers to deliver the programme sessions effectively?
- What did the RRW sessions look like in practice? How were the sessions structured? What rehearsal room techniques did teachers use and how? What learning opportunities did this provide the pupils with?

Given the unscripted and relatively unstructured nature of the programme, this research question will focus on understanding what RRW sessions looked like in practice and whether they provided the intended learning opportunities for pupils as outlined in the ToC. Variation will also be considered in terms of when and where the sessions are delivered, and how the sessions are incorporated into the class schedule and curriculum. Understanding more about what the sessions look like in practice will provide further detail for the ToC and inform understanding of the mechanisms of change that may be at play. We will also look to understand the ways in which teachers did (or did not) incorporate writing into the sessions and/or related the sessions back to writing, as this is required by the primary causal pathway in the ToC but not explicitly built into the programme materials. Together, this will provide important context for interpreting the Impact analysis findings, in terms of whether the sessions that were delivered met the objectives of the programme training and, hence, whether the findings represent the true effects of implementing the programme theory in practice.

Context & moderators

> **IPE RQ2:** *What key moderators and contextual factors were perceived to influence how effectively RRW was implemented?*

- What were the key challenges and facilitators to successfully implementing the programme?
- What factors relating to the school context and/or teacher characteristics were perceived to influence the effectiveness of programme implementation?
- What (if any) challenges or facilitators were observed that would be relevant to further scale-up of the programme in future?

This research question is concerned with understanding how implementation of the programme may have varied depending on the context in which it was delivered and/or the characteristics of the teacher delivering it. Challenges and facilitators relating to training attendance and

engagement will likewise be explored, including the logistics of the required in-person training attendance. Factors relating to implementation, as explored under IPE RQ1 will likewise be considered in relation to how they may influence any outcomes observed. The list of likely factors is provided in the ToC under 'Moderators'.

> **IPE RQ3:** *What (if any) barriers did disadvantaged pupils face specifically? What facilitators were seen to counter these?*

- To what extent were pupils eligible for FSM and/or with SEN able to benefit from the programme? What, if any, factors were perceived to affect this?

This research question will provide important context for interpreting the findings from the Impact analysis for RQ4, RQ10b and RQ11b. Understanding the experience of disadvantaged pupils is particularly important given both the place that Shakespeare holds in the English National Curriculum in relation to privilege pupil experiences,[5] and the positive feedback from teachers in the previous trial about the inclusive and accessible nature of RRW (McCulloch and Collins, 2023).

Responsiveness

> **IPE RQ4:** *How did (a) teachers and (b) pupils respond to the programme?*

The RRW training is intended to provide teachers with a toolkit for delivery in their classrooms; teachers are not simply able to rely on a script or pre-specified plan. Therefore, the extent to which teachers actively participate in the training (as a mechanism to gain embodied experience of the activities they may subsequently deliver) will be part of the IPE. Whether or not teachers enjoy using the techniques they have been taught may also influence the number and quality of sessions delivered for the same reason.

The RRW sessions are intended to be interactive, and so we will explore the extent to which pupils join in with the activities, and how teachers feel the pupils in their classes respond to the activities delivered. The programme activities and outputs centre on active and embodied pupil participation, so to provide implementation context for the Impact analysis findings it is important to understand not just fidelity in terms of what kinds of activities teachers used in the sessions (covered by IPE RQ1) but also what pupils did in response. We will explore whether pupil responsiveness varies between individuals and between schools, and what factors may be driving this. Pupil enjoyment of the sessions will be qualitatively gauged to help inform our interpretation of any change in attitudes towards reading that may be observed in the impact analysis (RQ5 and RQ6).

Perceived impact

> **IPE RQ5**: *What was the perceived impact of the programme for:*
>
> *a) Teachers*

[5] See, for example, Ward and Connolly, 2008; Purewal, 2017.

b) Pupils

c) Disadvantaged pupils

- To what extent was the programme perceived to improve teacher self-efficacy and/or job satisfaction?
- To what extent did pupils (a) enjoy and (b) feel motivated to write in response to the programme sessions?
- To what extent was the programme perceived to improve pupil writing self-efficacy, enjoyment of writing and/or writing attainment? To what extent were these outcomes observed for disadvantaged pupils specifically?
- To what extent did the programme appear to support pupils to grow their confidence around using spoken language?
- What (if any) other positive outcomes were perceived for teachers and/or pupils?
- What (if any) negative unintended consequences were reported?

This research question will complement the findings of the Impact analysis by looking to provide a more nuanced picture of programme impact as perceived by participants. In addition, it will seek to identify any additional outcomes that may have been observed by participants but were not accounted for in the pre-specified Impact analysis. This covers both academic and non-academic (e.g. cultural, socio-emotional) outcomes. Impact on teacher job satisfaction will be investigated here, as well as any effects for pupil feelings around communicating with other people (as a proxy for confidence around using spoken language). These outcomes have been theorised as likely benefits for programme participants but were not included in the Impact analysis as they were not seen to relate directly to the primary outcome of increased pupil writing ability.

This research question will also look to uncover any potential negative unintended consequences that may result from the programme. These may include learning displacement as a result of the loss of the learning time that the sessions replaced. Similarly, teacher absence to attend the training may result in learning loss (Benhenda, 2022). Potential programme drawbacks such as impact on teacher workload will also be investigated.

Programme differentiation

> **IPE RQ6:** *What was business as usual (BAU) and to what extent did it differ from lessons delivered as part of the RRW programme?*

- What did business as usual look like in all schools prior to the start of the trial?
- What did business as usual look like in control schools during the trial?
- To what extent did implementation of the RRW programme differ from business as usual in both intervention and control schools?

This question will explore the extent to which practice in control and treatment schools differed and, hence, the extent to which any difference in outcomes may be attributable to a programme effect – or may represent an underestimate of the true effect. We will look, in particular, at the teaching of Shakespeare and the use of drama-based and/or embodied learning in control schools during the trial.

Research methods

The IPE employs a range of mixed-methods data collection activities to provide data from different perspectives that can be triangulated to form a richer picture of the process of implementation. The IPE data collection activities occur in a staggered manner over the course of the trial, which will enable each data collection instrument to be informed by the context and emerging findings of previous activities (see Table 3).

All data collection instruments will be informed by the programme logic model, TIDieR framework and IPE research questions, in addition to findings drawn from the data already collected.

Table 3: Sequencing of the IPE data collection activities

<table>
<tr><th>Before randomisation</th><th colspan="2">During training and delivery (interleaved)</th><th>After delivery</th><th>Longitudinal</th></tr>
<tr><td rowspan="2">Baseline BAU teacher survey</td><td>Semi-structured observations of training days (x3)</td><td rowspan="5">Case study observations, teacher interviews & pupil focus groups (x8)</td><td>Endpoint BAU teacher survey</td><td rowspan="5">Longitudinal teacher survey</td></tr>
<tr><td>Focus groups with RSC trainers (x2)</td><td>Endpoint intervention teacher survey</td></tr>
<tr><td rowspan="3">Baseline pupil survey</td><td rowspan="2">Semi-structured review of training discussion outputs</td><td>Endpoint pupil Survey</td></tr>
<tr><td>Case study session trackers</td></tr>
<tr><td>Training completion data</td><td>Session Delivery Logs</td></tr>
</table>

An overview of the methods and data collection activities that will be used to respond to each of the IPE dimensions and research questions is provided in Table 4 below. Further detail about each activity is provided in the main text that follows.

Table 4: IPE methods overview

IPE dimension	IPE RQ addressed	Research methods	Data collection methods	Sample size and sampling criteria	Data analysis methods
Fidelity, adaptation & reach	IPE RQ1	Qualitative methods	Semi-structured observations	3 x training days; 8 x programme sessions	Thematic analysis
			Focus groups	2 x groups of 4 trainers; 8 x groups of 6 pupils	
			Interviews	8 x teachers	
			Semi-structured reviews	8 x session trackers; 20 x training discussion outputs	
		Quantitative methods	Online surveys	BAU surveys x2 (all teachers – baseline, control teachers - endpoints); Teacher survey (intervention teachers - endpoint)	Descriptive statistics
			Monitoring data	Session delivery logs / session trackers (intervention teachers); training attendance data (intervention teachers)	
Context, moderators & dosage	IPE RQ2	Qualitative methods	Semi-structured observations	3 x training days; 8 x programme sessions	Thematic analysis
			Focus groups	2 x groups of 4 trainers; 8 x groups of 6 pupils	

			Interviews	8 x teachers	
			Semi-structured reviews	8 x session trackers; 20 x training discussion outputs	
		Quantitative methods	Online surveys	Teacher survey (intervention teachers - endpoint)	Descriptive statistics
			Monitoring data	Session delivery logs / session trackers (intervention teachers); training attendance data (intervention teachers)	
	IPE RQ3	Qualitative methods	Semi-structured observations	3 x training days; 8 x programme sessions	Thematic analysis
			Focus groups	8 x groups of 6 pupils	
			Interviews	8 x teachers	
			Semi-structured reviews	8 x session trackers; 20 x training discussion outputs	
		Quantitative methods	Online surveys	Teacher survey (intervention teachers - endpoint)	Descriptive statistics
Responsiveness	IPE RQ4	Qualitative methods	Semi-structured observations	3 x training days; 8 x programme sessions	Thematic analysis
			Focus groups	2 x groups of 4 trainers; 8 x groups of 6 pupils	
			Interviews	8 x teachers	

			Semi-structured reviews	8 x session trackers; 20 x training discussion outputs	
		Quantitative methods	Online surveys	Teacher survey (intervention teachers - endpoint)	Descriptive statistics
Perceived impact	IPE RQ5	Qualitative methods	Focus groups	8 x groups of 6 pupils	Thematic analysis
			Interviews	8 x teachers	
			Semi-structured reviews	8 x session trackers; 20 x training discussion outputs	
		Quantitative methods	Online surveys	Teacher survey (intervention teachers - endpoint); Longitudinal teacher survey (intervention teachers – longitudinal)	Descriptive statistics; statistical testing (pupil feelings about communicating with other people, teacher job satisfaction)
			Paper surveys	Pupil survey (all – baseline and endpoint)	
Programme differentiation	IPE RQ6	Quantitative methods	Online surveys	BAU survey (all teachers – baseline, control teachers – endpoint)	Descriptive statistics

1. Training data collection

Qualitative data collected through observations of the training sessions, reviews of the training discussion outputs and focus groups with the trainers will be used to understand what the training looked like in practice and how the teachers responded to it. This data will inform the development of the subsequent data collection instruments and provide important context for later data collection activities, particularly for determining fidelity during the case study observations.

Semi-structured observations of training days

Three of the five in-person training days will be observed by an NFER researcher using a semi-structured observation tool. In line with the minimum attendance requirements for the training, an NFER researcher will observe the first two training days (including the evening performance) and the final training day (Days 1, 2 and 5). We will aim to observe each pair of trainers for at least one day to get a sense of any variation in the style or content of delivery that may occur. The observation tool will focus on gathering descriptive accounts of what took place rather than any kind of quantitative monitoring. This is with the aim of obtaining as rich a picture as possible of what are likely to be complex and dynamic sessions. The observation will look at both what is delivered and how the teachers respond. We will also record the reflections and feedback shared by teachers during trainer-led discussions, particularly in terms of challenges or facilitators they may have encountered.

Online focus groups with RSC trainers

The four RSC trainers participating in the trial will be asked to take part in an online focus group in November 2025 after the first block of training, and again in April 2026, after the final training day. Each focus group will last around 45 minutes and take place via Microsoft Teams. The sessions will be facilitated by an NFER researcher. Video recordings of the sessions will be taken with participants' permission to allow for more accurate records of the conversation to be made. The trainers will be asked about the training, whether it is going as intended, how the teachers are responding to it and any reflections on the insights teachers have shared – particularly in terms of challenges and facilitators for implementation. This information will provide context for the training session observations and inform design of the data collection tools for subsequent IPE activities.

Training discussion outputs

During the training sessions, the trainers collect teacher reflections in the form of sticky notes. Photographs of the sticky note boards from both cohorts for each of the three training blocks will be shared with the NFER team and qualitatively reviewed using a semi-structured tool closely aligned to the training observation tool to code teachers' experiences of the training and implementation process, including challenges and adaptations.

2. Case studies

We will carry out case study data collection activities with eight intervention schools to gather in-depth information about the experience of participating in the programme in relation to each of the IPE dimensions. Given its small size, this sample is not intended to be representative of the study population but to provide in-depth insights into programme implementation in a range of contexts. For this reason, we will seek to include a range of school characteristics (school sizes, class sizes and regions) as well as socio-economic circumstances.[6]

Schools identified by the purposive sampling will be invited to participate in the case study activities, with the option to decline. Schools that are more engaged with the programme may be more likely to agree to be a case study school, introducing some selection bias so we will take this into account when reporting the case study data. We will also be mindful of this when reporting the pupil focus groups as the selection of pupils by teachers (see below) is likely to also have some selection bias in terms of which pupils they put forward.

Only one class per school will participate in the case study activities and we will clearly communicate what will be expected of the school before confirming their participation. We will work closely with the key contact at the school to confirm the timings of each data collection activity in advance of the visit and to ensure this disrupts the class schedule as little as possible. Case study schools will be identified in the autumn term 2025 and visits will take place between January and April 2026.

Case study schools will receive a 'thank you' payment of £150 to encourage participation from schools regardless of their level of engagement in the programme itself. The participating teacher will receive a voucher of £20 in recognition of the additional time they give to the evaluation.

Triangulation of the observation data, pupil feedback and teacher interviews will allow us to build a picture of implementation in each school, any challenges and/or adaptations involved, the impact it is perceived to be having and the extent to which it is being implemented as intended.

Case study schools will also be asked to provide more detailed information about the cost of participating in the programme via the Cost Pro Forma (see 'Cost evaluation' section below).

Semi-structured session observation

The NFER researcher will use the case study visit as an opportunity to observe delivery of an RRW session. The researcher will aim to minimise disruption to the usual flow and dynamic of the session. A semi-structured observation tool will be developed based on the ToC, TIDieR framework and training data with the aim of collecting qualitative data about what happens during the session and how the pupils respond to it, as well as any relevant contextual factors such as the space in which it is delivered.

[6] School-level socio-economic circumstances will be assessed through the proxy measure of the percentage of pupils eligible for FSM recorded in the Get Information about Schools (GIAS) database (Department for Education, 2025b).

Pupil focus groups

Following the observation, six pupils will participate in an in-person focus group to share their views on the session the researcher observed. The six children will be selected by the teacher and written opt-in parental consent obtained in advance. The teacher will be asked to include pupils with different levels of ability and disadvantage (e.g. FSM eligibility) when selecting the group.

The children will be asked how they felt about the session, what they liked more and less about it, and how they feel about writing more generally. The focus groups will last around 20 minutes and will involve interactive, age-appropriate activities as a means of helping the pupils to feel comfortable about sharing their experiences.

All researchers involved in these data collection activities will have an up-to-date enhanced DBS check and safeguarding training and follow the school and NFER's safeguarding protocols in all instances. A member of school staff will be present at all times, though not directly participating in the session. While this may influence pupil responses, this is important for safeguarding purposes. Moreover, children of this age are likely to feel more comfortable in the presence of a familiar adult, which would support openness and engagement.

Teacher interviews

The teacher will be asked to participate in an interview either in-person on the day of the visit (after the observation and pupil focus group) or remotely via Microsoft Teams at another time convenient for them. The interview will last approximately 45 minutes and will be audio recorded (or video recorded if remote) with the participant's permission. Teachers will be invited to refer to their Reflective Journal during the interview if they would find this helpful. The interviews will ask about the teacher's experience of the training and of delivering RRW sessions with their class, including challenges, facilitators and how they feel their pupils have responded to it.

Session tracker

In addition to the visit, case study teachers will be asked to complete a Session Tracker, where they provide a high-level summary of each of the sessions they deliver. This would constitute one sentence on the session objective and a high-level summary of the session structure (e.g. warm up, choral reading, tableaux, cool down, writing task), alongside the date the session was delivered. The tracker will be in the form of a Microsoft Word document shared with the teacher via NFER's secure online portal. The tracker will be shared with the teacher in the autumn term 2025, as soon as their participation in the case study activities have been confirmed, so that they can start using with their sessions as soon as possible.

It is possible that teachers using the session tracker may be prompted to take a more structured approach to planning the RRW sessions compared to their peers. As a result, data from the session trackers will not be treated as representative of what occurred in all trial schools, but as examples of how the structure and approaches of each session may differ (or not) both across the implementation period and between teachers (even within the small sample).

Case study teachers will *not* be asked to complete the session delivery log (see below) due to the substantial overlap between the two instruments. Data about when and how many sessions were delivered will instead be extracted from the session tracker by the evaluation team for case study schools.

3. Surveys

The BAU, Teacher and Pupil surveys will collect data to inform analysis for the both the Impact and IPE. Only the data relating to the IPE is covered below. Please see 'Impact evaluation design' for more information about the Impact data.

BAU teacher surveys

Baseline: All teachers participating in the trial will be asked to complete an online survey prior to randomisation (September 2025) about their usual practice. The survey will take around 10 minutes to complete and will ask about the teacher's approach to teaching writing, their experience with dialogic, drama-based and embodied teaching strategies and any recent writing-related CPD. Teachers will also be asked whether their school already teaches Shakespeare and their confidence to do so.

Teachers will also be asked five items relating to job satisfaction from the Trends in International Mathematics and Science Study (TIMSS) 2019 Teacher Questionnaire (IEA, 2018). This is in addition to the four items relating to teacher efficacy for improving pupils' writing performance used for the Impact analysis (see 'Impact evaluation design' for more information).

Endpoint: Teachers allocated to the control group will be asked to complete a similar survey at endpoint (summer 2026) in relation to the period over which the trial occurred.

Intervention Group Teacher Survey

Endpoint: Teachers allocated to the intervention group will be asked to complete an online survey at the end of the programme delivery period (summer 2026). The survey will take around 10-15 minutes to complete and will involve routing to ensure respondents are only shown questions relevant to them. The survey will ask about the teacher's experience of the training, programme implementation and perceived outcomes for themselves and their pupils. The survey design will be informed by findings from the case study data collection activities, with the aim of understanding the prevalence of some of the perceptions and occurrences that emerged.

The Teacher Survey will also include the items from the TIMSS 2019 Teacher Questionnaire and Teacher Efficacy Scale for Writing from the BAU surveys.

The BAU and Teacher surveys will be conducted via the survey platform Tivian. Each participant will receive a unique link.

Pupil surveys

For the purpose of the IPE, all pupils will be asked at baseline and endpoint to answer questions relating to how they feel about communicating with others. These questions will be drawn from

the Fear of Communication measure used with UK pupils in years 5 and 6 for the effectiveness trial of Children's University (Morton, Dirie and Palak, 2023). The measure consists of eight items relating to how pupils feel about communicating with other people, adapted from the Personal Report of Communication Fear (McCroskey *et al.*, 1981). Pupils will also be asked items relating to ideation and self-regulation from the Self-Efficacy for Writing Scale and items from the Liking Writing Scale (Bruning *et al.*, 2013) for the Impact analysis (see 'Impact evaluation design' for more detail).

Paper-based surveys will be administered to pupils by the class teacher (at baseline) or NFER Test Administrator (at endpoint) immediately before the NMM writing assessment (with a short break in between). In order to minimise the risk of confounding due to reading ability levels, teachers and Test Administrators will read out the survey questions and age-appropriate response options. While visual response scales were considered, it was not possible for these to be sufficiently aligned with the requirements of the pre-defined measure. The age-appropriate response scales will be tested with a small number of children in advance of the trial administration. Teachers and Test Administrators will be provided with instructions and guidelines for the kinds of clarifications that they will be permitted to make. The survey will take around 10 minutes to complete.

4. Monitoring data

Session delivery logs

All teachers allocated to the intervention group (with the exception of case study teachers) will be asked to complete a session delivery log to record when and how many hours of RRW sessions they deliver. The log will take the form of a Microsoft Excel spreadsheet that will be shared with schools via NFER's secure online portal. The logs will be shared with teachers prior to the start of session delivery. This data will be used in the compliance analysis (see 'Impact evaluation design') as well us to provide a descriptive understanding of variation in the number and frequency of sessions delivered.

Training attendance data

The RSC team will provide NFER with a spreadsheet outlining which of the training days each teacher allocated to the intervention group attended. As with the session delivery logs, this data will be used in the compliance analysis (see 'Impact evaluation design'), as well as to provide an overview of levels of fidelity given the importance of the training in the programme structure.

Reasons for withdrawal

Any schools that ask to withdraw from the trial and/or programme delivery will be asked to respond via email to a single multiple-choice question indicating the reason for their withdrawal.

5. Longitudinal IPE

The longitudinal Impact analysis (see 'Impact evaluation design' section) of pupil outcomes will be complemented by a light-touch IPE concerned with teacher outcomes. At the follow up

timepoint, the teachers are unlikely to be teaching the Year 6 pupils in the analysis, so any contrasts drawn between the datasets will be descriptive in nature only.

One year after the end of the programme (summer 2027), teachers that had been allocated to the intervention group will be asked to complete an online survey. The survey will ask about any continued use of RR techniques in the teacher's practice in relation to writing or other subject, as well as any spreading of these techniques to other classes. It will take around 5 minutes to complete.

We are aware that the sample of respondents will be highly skewed towards teachers who were more engaged with the programme and thus more likely to have integrated the techniques into their ongoing practice. This limitation will be clearly accounted for when reporting these findings.

Analysis

All IPE analysis will be carried out in line with the EEF's guidance (EEF, 2022a).

Qualitative data – interviews, observations, focus groups and reviews

Observation and review data will be treated qualitatively to provide a clear narrative of what the programme looks like in practice and the key variables that influence the nature of its implementation. Notes from the observations, reviews, interviews and focus groups will be written up as intelligent verbatim transcripts[7] and uploaded to the qualitative data analysis software MAXQDA (VERBI Software, 2021). The data will then be analysed using thematic analysis to identify and analyse both manifest and latent patterns of meaning (Joffe, 2012). This will involve an initial process of deductive coding from a pre-determined coding frame based on the ToC, IPE dimensions and research questions, followed by a more detailed inductive coding process to draw out the key themes emerging from the texts.

The MAXQDA 'variables' function will support within-case analysis by making it possible to look at all the data from a case study for each code in one place (MAXQDA, 2022). This will help us to triangulate the different data sources and better understand the context in which particular issues or perspectives emerge. At the same time, inter-case study analysis will be facilitated by the option to view all data per code across one or more data source types.

At least two researchers will be involved in the coding process to reduce the risk of bias. A common approach will be discussed and agreed upon by the coding team in advance, and an initial subset of the data will be double-coded and discussed to support inter-coder reliability. This will be facilitated by the use of MAXQDA's Intercoder Agreement function, which facilitates comparison of the independent coding of the same text performed by two different people (MAXQDA, 2025). Further cross-checking will occur whenever necessary, with questions and uncertainties addressed collaboratively by the coding team.

[7] Intelligent verbatim transcription excludes fillers and redundancies that do not add meaning to the content to make the text more 'readable' (McMullin, 2023).

Quantitative data – surveys and monitoring data

Survey response data will be exported from Tivian and quality assured prior to its analysis, with the data from each source stored in a separate file.

The analysis plan for the survey and monitoring data will be designed by the team statistician in consultation with the Project & IPE Lead and Project Director. After data collection the statistician will then carry out the analysis using R, with tables and charts produced in Microsoft Excel. All codes and outputs will be checked by another experienced member of NFER's Centre for Statistics. The analysis will cover a range of descriptive statistics as required to address the IPE research questions, including a small number of cross-tabulations of key variables to better understand the relationship between them. These variables will be selected in advance as part of the analysis plan. For example, we may look to understand if and how both staff and pupil survey responses vary according to school-level disadvantage (through the proxy of the proportion of pupils eligible for FSM).

In addition, we will use linear multi-level modelling to statistically test for differences between intervention and control groups at endpoint in relation to pupil-reported feelings about communicating with other people, controlling for baseline scores as a covariate. The outcome for this model will be the sum of the eight items, where negatively worded questions are reverse coded. For the teacher-reported job satisfaction measure, in line with the TIMSS methodology (Mullis *et al.*, 2020), the five items will be additively combined and predefined cut scores will be used to classify teachers as Very Satisfied, Somewhat Satisfied or Less than Satisfied. We will use an ordinal multi-level model to statistically test for differences in teacher-reported job satisfaction between intervention and control groups at endpoint, controlling for baseline scores as a covariate and accounting for clustering within schools.

Triangulation of qualitative and quantitative data

The data collection tools across both the qualitative and quantitative methods will be designed to facilitate triangulation and complementary analysis. For example, variations of certain questions will be asked in both the teacher interviews and Teacher Survey, to allow for both an in-depth understanding of that particular dimension as well as the prevalence of the various perspectives that emerge. Considerations relevant to future data collection tools that emerge from the data collection activities will be logged in real-time and referred to during the instrument design process.

We will also develop an integrated analysis framework to guide our analysis and reporting for each of the IPE dimensions and research questions. An IPE analysis workshop will be carried out prior to report writing to cross-check interpretations of the data and to develop an outline of the key points to be covered under each of the research questions. A subsequent analysis workshop with the broader evaluation team will bring the IPE and impact analysis findings together to facilitate reflection on any insights that the IPE may bring to the impacts that may or may not have been observed.

Cost evaluation design

There is not currently any cost evaluation information available regarding participation in the RRW programme. As a result, we will monitor the cost in terms of time and expenditure that intervention schools incurred from implementing the programme as part of the trial. We will collect information on the pre-requisite, set-up and ongoing costs to schools, in accordance with the EEF cost evaluation guidance (EEF, 2023).

As much of the cost information as possible will be collected directly from the RSC to minimise burden on schools. Where data collection from schools is required, it will be embedded within the IPE activities (surveys and case studies) for the same reason and to enable triangulation from various sources. The main cost evaluation data collection tool will be the cost pro forma that will be administered to the case study schools. This will be a short questionnaire in a Word document about the costs incurred as part of the training and programme implementation. We will ask that the teacher liaise with members of their SLT when completing it to provide the most complete information possible. A digital copy of the cost pro forma will be shared via NFER's secure online portal ahead of the case study visit and the teacher will have the opportunity during the visit to ask any questions about completing it. Teachers will be asked to return the completed cost pro forma at the end of the delivery period (June 2026). The Teacher Survey will also include low-burden cost questions that will provide context on how representative of the intervention schools the case study sample may be seen to be.

As the cost of attending the training is subsidised for this trial, the RSC will have access to all the data relating to the costs incurred in relation to the training. As a result, we will collect the following data about the schools from the RSC:

- Average number and cost of overnight stays per school to attend the training
- Average cost per school of travel to the training

The RSC will also be able to provide information about the length of the training sessions and the likely market price.

In addition, we will ask schools about:

- Number of hours of classroom/teaching cover required to attend the training and nature of the cover provided
- Time spent travelling to and from the training
- Any other costs for attending the training *not* covered by the trial
- Any material costs for the sessions
- Time taken to prepare for each of the sessions
- Duration of the sessions

- Time spent on any other related activities

We will also look to understand the extent to which time spent on the programme was in addition to or instead of the teacher' usual workload and, if the former, whether this additional time was paid.

The cost evaluation analysis will be carried out in line with the EEF's guidance (EEF, 2023). Each cost to the schools will be estimated per year over a projected three-year period. We will differentiate between pre-requisite costs, start-up costs and recurring costs. Ongoing delivery costs in years 2 and 3 will either be reduced to zero (for fixed cost) or we will make an informed decision about whether they are likely to change over time. Time and cost estimates will be reported in terms of means and ranges. Once we have established the cost per-school-per year, we will divide this by the number of intervention pupils per school to estimate the cost per-pupil-per-year.

Sensitivity analyses will be carried out for any costs (including staff time) that demonstrate meaningful variability. This is likely to include the cost of transport to the training and time taken to prepare for each of the sessions.

Ethics and registration

This evaluation will be conducted in accordance with NFER's Code of Practice. All of NFER's projects abide by its Code of Practice, which is in line with the Codes of Practice from BERA (the British Educational Research Association), MRA (the Market Research Association) and SRA (the Social Research Association), among others. NFER is committed to the highest ethical standards in all of its activities and ethical considerations are embedded in its detailed quality assurance processes. Every project is assessed against the NFER Code of Practice at proposal stage, with ethical approval a requirement of proceeding with the bid. Any significant updates to the project after this point are submitted to the Committee for further approval as needed.

Each participating school's headteacher will provide their agreement on behalf of the school to participate in the trial by signing the Memorandum of Understanding (MoU), which outlines the responsibilities of all parties involved in the trial.

NFER will share a parent letter and withdrawal form with schools to be sent to parents/carers of all pupils in participating classes. Through the withdrawal form, parents/carers will have the opportunity to withdraw their child from the evaluation and associated data processing at any stage of the trial.

A separate opt-in consent process will be used for the pupil focus groups and will only apply to those selected to participate. Given that pupils participating in this study are only 9 to 10-years-old, we cannot assume that all pupils will have the capacity to provide fully informed consent to participate. We will therefore provide parents/carers with a written information sheet about the focus groups which will contain full details about the focus group and what their child will be asked to do. Parents/carers will then be asked to provide written opt-in consent of their

willingness for their child to be invited to participate in the focus group, by returning a consent form to the school, who will then pass this information on to the research team.

Pupil participation in the focus groups is voluntary, therefore even if a parent/carer has given consent for their child to participant, their child can still choose not to take part. The visiting NFER research will read age-appropriate information about the focus groups to the pupils at the beginning of the focus group to ensure pupils understand it and have the chance to ask any questions. If, at this point, a pupil decides that they would prefer not to participate, then they will be able to return to their class. Prior to beginning the focus group, the researchers will agree some ground rules for the group with the pupils and have a discussion with them about the types of scenarios (that is, safeguarding concerns) in which we would need to break confidentiality, to ensure they fully understand what this means. A school staff member will also be present at all times.

All case study researchers, Test Administrators, judges, transcribers and coders will receive training in identifying and responding to any potential disclosures or other safeguarding concerns.

In addition, we will conduct a safeguarding check prior to the judging process[8] (for the written tasks that form the primary outcome). This process will take place twice, at baseline and at endpoint, once the completed scripts have been received at NFER. A team of trained safeguarding (SG) reviewers will read all completed assessments/scripts and any instances of concern will be flagged for further review. Flagged assessments will be reviewed by the designated SG team (2/3 core members of the project team including the Operations Lead and Project Director/Leader) within 24 hours and categorised and followed up as shown in the table below. If necessary, the SG team will consult with NFER's Designated Safeguarding Lead (DSL) for a second opinion.

Category	Action
a) no concern	No follow up required, case documented and closed
b) no obvious concern that meets the safeguarding threshold, but the team feels it would still be beneficial to share with school in case they feel there is a need to follow up (e.g. concerning content but no direct link to writer)	School contacted by phone/email, script shared securely if requested by school. Case documented and closed.
c) safeguarding concern identified	Immediate follow up required. The designated SG team member will phone the

[8] Originally we had planned to complete safeguarding checks during the judging process but due to the time lag between pupils completing the assessment and the judging, it was felt that this was too long to leave such checks (this would have been a particular issue for the baseline writing assessment as this is not being judged until endpoint, i.e. almost a year after pupils complete it).

	school within 24 hours to share details of the SG concern with the school's DSL (or deputy) or an SLT member. Script shared securely if requested by school. Case documented including follow up actions (if applicable) and closed once school has taken on the case.

Any safeguarding concerns will be raised with the Project Director and/or Lead within 24 hours of being identified. The Project Director and/or Lead will review all concerns raised with NFER's Designated Safeguarding Lead and escalate to the school for further action where appropriate. NFER will notify the EEF of any safeguarding incidents as soon as possible.

Interviewees (e.g. school staff) will be provided with information about the research and how we use their data before our visit and informed consent will be obtained from interviewees at the start of the interview. If an individual staff member within a case study school does not wish to participate in the data collection they can choose to decline.

The trial will be designed, conducted and reported to CONSORT standards. It will be registered in the Open Science Framework (OSF) Registry once the protocol has been finalised. The OSF number will be added to this document as soon as it becomes available. The registry will be updated will the trial outcome upon its completion.

Data protection

All data gathered during the trial will be held in accordance with the data protection framework created by the Data Protection Act 2018 and the General Data Protection Regulation 2016/679 and will be treated in the strictest confidence by the NFER, the RSC and the EEF. No individual or school will be identified in any report.

NFER is the data controller for evaluation and will make decisions about how and what personal data is used in accordance with the objectives of the study set by EEF. The RSC is the data processor for the evaluation and data controller for the programme delivery.

The legal basis for processing personal data is covered by GDPR Article 6 (1) (f):

Legitimate interests: The processing is necessary for your (or a third party's) legitimate interests unless there is a good reason to protect the individual's personal data, which overrides those legitimate interests.

A legitimate interest assessment has been undertaken. The evaluation fulfils one of NFER's core business purposes (undertaking research, evaluation, and information activities). It has broader societal benefits and will contribute to improving the lives of learners by providing evidence about the impact of teaching techniques used in the classroom – in this case, the teaching of writing in the primary school classroom, and the relative effectiveness of the RRW programme specifically.

We do not believe this processing will cause damage or distress to the pupils. The outcomes of the evaluation will not result in the creation of measures or decisions being made about individual pupils.

NFER and the RSC will sign a Data Sharing Agreement (DSA) to govern the collection and sharing of personal data during this trial. This agreement includes a description of the nature of the data being collected and how it will be shared, stored, protected and reported by each party. In addition, the RSC will provide a memorandum of understanding to schools, explaining the nature of the data being requested of schools, teachers and pupils, how it will be collected, and how it will be passed to and shared with NFER. Two separate Privacy Notices are available: one for schools and another one for parents.[9] All personal data will be shared via secure, password-protected data sharing portals.

All NFER staff visiting schools will have up-to-date DBS checks. All data gathered during interviews will be stored securely. No names of individuals will be used in any report arising from this work.

Within three months of the end of project, NFER will send school and pupil data to EEF's data archive partner. At this point, EEF's data archive partner will keep a copy of the data and EEF will become the Data Controller. NFER will retain personal data for one year after report publication in case there are any queries about the report. One year after the report publication, all personal data will be securely deleted.

NFER will not store or transfer any data outside of the UK. When we use Tivian to administer online surveys, data is stored in the EU.

[9] Both privacy notices are available here: https://www.nfer.ac.uk/for-schools/participate-in-research/participate-in-research-projects/efficacy-trial-of-the-rehearsal-room-writing-programme/

Personnel

Table 5: Project team

Name	Organisation	Role and Responsibilities
Evaluation team		
Helen Poet	NFER	Project Director – responsible for overall delivery of the evaluation to agreed specifications and overseeing the integration of the impact and IPE. She will also be responsible for strategic oversight and the quality of the outputs.
Lillian Flemons	NFER	Trial Manager & IPE Lead – day-to-day management of the trial, and design and delivery of the IPE.
Ruth Staunton	NFER	Trial Statistician – leading on the impact design and statistical analysis.
Eleanor Bradley	NFER	IPE Researcher – IPE data collection
Sophie Ainsby	NFER	Research Operations Lead – responsible for leadership and strategy around data collection and school communications
Lydia Wallis	NFER	Operations Manager – day-to-day operations, including preparation of recruitment documents, coordinating data collection and point of contact for schools participating in the trial
Professor Annabel Watson	University of Exeter	Writing Expert – advising on outcome measures and providing knowledge of writing pedagogy
Professor Phil Durrant	University of Exeter	Linguistics Expert – conducting analysis on vocabulary diversity and noun phrase complexity
Delivery team		
Jacqui O'Hanlon	RSC	Project Director – leading overall delivery of the programme and responsible for strategic oversight and quality control
Matthew Collins	RSC	Programme Lead – day to day operations, including project development, school recruitment and programme delivery

Lynsey McCulloch	RSC	Programme Lead – day to day operations, including project development, school recruitment and programme delivery
Professor Maggie Snowling	University of Oxford	Educational Psychology Expert – consultant to the delivery team for the project

Risks

Table 6: Project risks

No.	Risk	Risk Assessment		Mitigation/Counter Measures/Contingencies
		Likelihood	Impact	
1	Insufficient schools recruited to the trial	Medium	High	NFER to input into recruitment material and work closely with the RSC throughout the recruitment process. If required, NFER's experienced operations team will assist with recruitment through a separate grant agreement. NFER and the RSC will decide and monitor pre-agreed recruitment targets to identify any unfavourable trends early on to act quickly. NFER will take an efficient and flexible approach to teacher and pupil data collection to allow for the recruitment window to be open for as long as possible. Estimation of the MDES has considered possibility of recruitment target (183 schools) not being met. Some contingency is assumed.
2	School and pupil attrition from trial and primary analysis	Moderate	Significant	Schools sign up for the trial via a Memorandum of Understanding with a clear identification of requirements. NFER will provide clear initial and ongoing communication via one key contact per school explaining principles and expectations. We will keep them informed of upcoming activities, timelines and next steps and provide support on all activities to ensure that activities are completed. School burden at endpoint will be minimised through the use of NFER Test Administrators to administer endpoint tests, at a convenient time for the school.

				Estimation of the MDES includes contingency for 10% school-level and 15% pupil-level attrition.
3	Teacher dropout from deliver / turnover	Low	Moderate	Trial activities (with the exception of the Longitudinal Teacher Survey) will take place within a single academic year, reducing the risk of turnover. NFER will work closely with RSC and schools to understand contingency plans for teacher turnover in intervention schools. Reasons for any dropout that does occur will be captured wherever possible.
4	The programme is not well implemented	Low	Moderate	Intervention schools will be provided with clear information about expectations and principles for the trial. Teachers will be provided with in-depth programme training. The process evaluation will monitor implementation.
5	Difficulty in securing target response rates for IPE	Moderate	Moderate	NFER will clearly communicate to school the value of these activities for the research. NFER will carry out a strategic campaign of invitations, follow-ups and reminders via email and phone calls to reach schools without creating pressure or additional workload. NFER will be highly flexible about timings for the case study activities to best suit school schedules. The RSC will promote IPE participation in their own communications with schools. Online data collection (including remote interviews) will be offered wherever possible to minimise burden and maximise flexibility of timing for participants. Case study schools will receive a 'thank you' payment of £150 in recognition of their time. The teacher at these schools will also receive a £20 voucher.
6	Difficulty securing parental permission for pupil participation in focus groups	Low	Low	NFER will contact schools to identify case studies as soon as possible and communicate what this involves, including pupil focus groups. NFER will share parental permission forms with schools as soon as they have been confirmed as case studies.

				NFER will provide a reminder to the school in the lead up to the case study visit to secure permission for the pupil focus groups.
7	Schools do not complete session delivery log	Moderate	Moderate	During set-up we will establish the minimum data that needs to be collected in the log to minimise burden. NFER will work to design a manageable instrument and completion process. NFER's operations team will support schools with completion and be on hand to answer queries.
8	Changes to the project team due to sickness, absence or staff turnover	Moderate	Low	NFER has a large research department with numerous experienced researchers and research who could be redeployed. Clear and accurate project documentation would support continuity in the event of any team changes.
9	Administering the pupil survey and writing task one after the other affects performance in the assessment	Low	Low	The two tasks are to be administered on the same day to reduce burden and disruption for schools. Schools will be provided with guidance about the content of the survey and how to talk about this with their pupils. The survey is not solely about writing, it also ends with questions about how they feel about speaking in front of/to others Pupils will be given a break between the two activities. The administration instructions and process are the same across both groups so any impact of administration approach is expected to be the same.

Timeline

Table 7: Timeline

Dates	Activity	Staff responsible/ leading
November – December 2024	IDEA Workshop and set-up meetings	NFER & RSC
January - February 2025	Development of recruitment materials and privacy notices	NFER & RSC

February – mid-June 2025	School recruitment	RSC
Mid-June – July 2025	Sharing of school data with NFER Collection of teacher and pupil data from schools Protocol publication & trial registration	RSC NFER NFER & University of Exeter
August 2025	Preparation for baseline activities (surveys & assessments)	NFER
September 2025	Baseline survey (BAU) Baseline assessment & pupil survey (teacher-administered)	NFER
October 2025	Randomisation and allocations communicated to schools NPD application Selection of case study schools	NFER
November – mid June 2026	Training days & intervention delivery IPE data collection (training observations, trainer focus groups, case study visits) Session delivery log & session tracker data collection SAP publication	RSC NFER
Mid-June – July 2026	Endpoint surveys (teacher & BAU) Endpoint assessment & pupil survey (administered by NFER test administrators) 'Thank you' payments to case study schools	NFER
August 2026	Assessment judging Sample of scripts transcribed for RQ2/RQ3	NFER
September 2026	'Thank you' payments to control schools Data cleaned and shared with NPD team Assessment results communicated to schools	RSC NFER
October – February 2027	Data analysis and reporting (draft report due end of February)	NFER & University of Exeter
March – May 2027	Report review process	NFER & RSC
June 2027	Publication of the main report Updating of the OSF Longitudinal teacher survey	NFER

October 2027 – February 2028	Longitudinal analysis and reporting (draft addendum due to the EEF at the end of February)	NFER
March – May 2028	Review period for the addendum	NFER & RSC
June 2028	Publication of the addendum Data archiving & updating of the OSF	NFER

References

Benhenda, A. (2022) 'Absence, substitutability and productivity: Evidence from teachers', *Labour Economics*, 76, p. 102167. Available at: https://doi.org/10.1016/j.labeco.2022.102167.

Brezina, V., Hawtin, A. and McEnery, T. (2021) 'The Written British National Corpus 2014 – design and comparability', *Text & Talk*, 41(5–6), pp. 595–615. Available at: https://doi.org/10.1515/text-2020-0052.

Bruning, R., Dempsey, M., Kauffman, D., McKim, C. and Zumbrunn, S. (2013) 'Examining dimensions of self-efficacy for writing', *Journal of Educational Psychology*, 105(1), pp. 25–38. Available at: https://doi.org/10.1037/a0029692.

Camacho, A., Alves, R.A. and Boscolo, P. (2021) 'Writing Motivation in School: a Systematic Review of Empirical Research in the Early Twenty-First Century', *Educational Psychology Review*, 33(1), pp. 213–247. Available at: https://doi.org/10.1007/s10648-020-09530-4.

Clark, C., Picton, I. and Cole, A. (2024) 'Children and young people's writing in 2024'.

Clarkson, R. (2024) '"It's missing the heart of what writing is about": teachers' interpretations of writing assessment criteria', *British Educational Research Journal*, 50(1), pp. 134–161. Available at: https://doi.org/10.1002/berj.3916.

Department for Education (2013) 'The national curriculum in England - Key stages 1 and 2 framework document'. Available at: https://assets.publishing.service.gov.uk/media/5a81a9abe5274a2e8ab55319/PRIMARY_national_curriculum.pdf.

Department for Education (2024a) 'Curriculum and Assessment Review: Review Aims, Terms of Reference and Working Principles'. Available at: https://assets.publishing.service.gov.uk/media/66d196b7d107658faec7e3db/Curriculum_and_assessment_review_-_aims_terms_of_reference_and_working_principles.pdf.

Department for Education (2024b) *Schools, pupils and their characteristics, Academic year 2023/24*, *GOV.UK*. Available at: https://explore-education-statistics.service.gov.uk/find-statistics/school-pupils-and-their-characteristics (Accessed: 28 March 2025).

Department for Education (2025a) *Attainment by pupil characteristics, Data set from Key stage 2 attainment*, *GOV.UK*. Available at: https://explore-education-statistics.service.gov.uk/data-catalogue/data-set/952fe280-2275-41f4-999f-f57f2a224653 (Accessed: 11 March 2025).

Department for Education (2025b) *Get Information about Schools*, *GOV.UK*. Available at: https://get-information-schools.service.gov.uk/ (Accessed: 20 March 2025).

Department for Education (2025c) *Key stage 2 attainment, Academic year 2023/24*, *GOV.UK*. Available at: https://explore-education-statistics.service.gov.uk/find-statistics/key-stage-2-attainment (Accessed: 11 March 2025).

Department for Education (2025d) 'The writing framework'. Department for Education. Available at:

https://assets.publishing.service.gov.uk/media/686e7890fe1a249e937cbecb/The_writing_framework.pdf.

Durrant, P. and Brenchley, M. (2023) 'Development of Noun Phrase Complexity Across Genres in Children's Writing', *Applied Linguistics*, 44(2), pp. 239–264. Available at: https://doi.org/10.1093/applin/amac032.

Durrant, P., Brenchley, M. and McCallum, L. (2021) *Understanding development and proficiency in writing: quantitative corpus linguistic approaches*. Cambridge: Cambridge University Press.

Education Endowment Foundation (2021) *Arts participation*, *EEF*. Available at: https://educationendowmentfoundation.org.uk/education-evidence/teaching-learning-toolkit/arts-participation (Accessed: 11 March 2025).

Education Standards Research team (2012) *The research evidence on 'writing'*. London: Department for Education. Available at: https://assets.publishing.service.gov.uk/media/5a7b63a740f0b6425d592d3f/what_is_the_research_evidence_on_writing.pdf.

EEF (2022a) *Implementation and process evaluation guidance for EEF evaluations*. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/evaluation-design/EEF-IPE-Guidance-August-2022.pdf?v=1739885156 (Accessed: 16 April 2024).

EEF (2022b) *Statistical analysis guidance for EEF evaluations*. Available at: https://d2tic4wvo1iusb.cloudfront.net/documents/evaluation/evaluation-design/EEF-Analysis-Guidance-Website-Version-2022.14.11.pdf?v=1669318235 (Accessed: 21 October 2024).

EEF (2023) *Cost evaluation guidance for EEF evaluations*. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/evaluation-design/Cost-Evaluation-Guidance-Feb_2023.pdf?v=1739793629 (Accessed: 26 February 2025).

Gilbert, J. and Graham, S. (2010) 'Teaching Writing to Elementary Students in Grades 4-6: A National SUrvey', *The Elementary School Journal*, 110(4).

Graham, S., Harris, K.R., Kiuhara, S.A. and Fishman, E.J. (2017) 'The Relationship among Strategic Writing Behavior, Writing Motivation, and Writing Performance with Young, Developing Writers', *The Elementary School Journal* [Preprint]. Available at: https://doi.org/10.1086/693009.

Grima, G., Hooper, A., Mullins, L., Redmond, B., Sharott, P. and Ashworth, B. (2024) 'Understanding current practice and research priorities in teaching Writing'.

Harrison, M.G., King, R.B. and Wang, H. (2023) 'Satisfied teachers are good teachers: The association between teacher job satisfaction and instructional quality', *British Educational Research Journal*, 49(3), pp. 476–498. Available at: https://doi.org/10.1002/berj.3851.

IEA (2018) 'TIMSS 2019 Teacher Questionnaire - Grade 4'. IEA. Available at: https://timssandpirls.bc.edu/timss2019/questionnaires/index.html.

Joffe, H. (2012) 'Thematic Analysis', in D. Harper and A.R. Thompson (eds) *Qualitative research methods in mental health and psychotherapy: a guide for students and practitioners*. Chichester: John Wiley & Sons, Ltd., pp. 209–223.

Laurin, S. (2010) *The effect of story drama on children's writing skills*. masters. Concordia University. Available at: https://spectrum.library.concordia.ca/id/eprint/979297/ (Accessed: 11 March 2025).

Lee, B.K., Patall, E.A., Cawthon, S.W. and Steingut, R.R. (2015) 'The Effect of Drama-Based Pedagogy on PreK–16 Outcomes: A Meta-Analysis of Research From 1985 to 2012', *Review of Educational Research*, 85(1), pp. 3–49. Available at: https://doi.org/10.3102/0034654314540477.

Lopes, J. and Oliveira, C. (2020) 'Teacher and school determinants of teacher job satisfaction: a multilevel analysis', *School Effectiveness and School Improvement*, 31(4), pp. 641–659. Available at: https://doi.org/10.1080/09243453.2020.1764593.

Lyu, C. and Deng, S. (2024) 'Effectiveness of embodied learning on learning performance: A meta-analysis based on the cognitive load theory perspective', *Learning and Individual Differences*, 116, p. 102564. Available at: https://doi.org/10.1016/j.lindif.2024.102564.

Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. and McClosky, D. (2014) 'The Stanford CoreNLP Natural Language Processing Toolkit', in K. Bontcheva and J. Zhu (eds) *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. Baltimore, Maryland: Association for Computational Linguistics, pp. 55–60. Available at: https://doi.org/10.3115/v1/P14-5010.

MAXQDA (2022) *Document and code variables in MAXQDA*. Available at: https://www.maxqda.com/help-mx22/variables/document-code-variables-maxqda (Accessed: 17 February 2025).

MAXQDA (2025) *Intercoder agreement*. Available at: https://www.maxqda.com/help-mx22/teamwork/problem-intercoder-agreement-qualitative-research (Accessed: 17 February 2025).

McCulloch, L. and Collins, M. (2023) *Time to Act - randomised control trial: final report*. Stratford-upon-Avon: Royal Shakespeare Company. Available at: https://cdn2.rsc.org.uk/sitefinity/education-pdfs/articles-and-reports/time-to-act-report.pdf?sfvrsn=5aea8921_2.

McGlade, H. and Bergin, A. (2025) *What is needed from education policy to boost writing outcomes?* London: Education Policy Institute. Available at: https://epi.org.uk/wp-content/uploads/2025/03/EPI_OUP-Summary-Paper_Boosting-writing-outcomes_March25.pdf.

McMullin, C. (2023) 'Transcription and qualitative methods: implications for third sector research', *Voluntas*, 34(1), pp. 140–153. Available at: https://doi.org/10.1007/s11266-021-00400-3.

Morton, C., Dirie, A. and Palak, R. (2023) 'Evaluation of Children's University, effectiveness trial - Statistical Analysis Plan'. Education Endowment Foundation. Available at:

https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/Evaluation-of-CU-SAP-Final-version-April-2023_CLEAN.pdf?v=1750759429.

Mullis, I.V.S., Martin, M.O., Foy, P., Kelly, D.L. and Fishbein, B. (2020) *TIMSS 2019 International Results in Mathematics and Science - Classroom Contexts*, *IEA TIMSS & PIRLS*. Available at: https://timss2019.org/reports/classroom-contexts/index.html#classroom-teacher-prep.

Pajares, F. (2003) 'Self-Efficacy Beliefs, Motivation, and Achievement in Writing: A Review of the Literature', *Reading & Writing Quarterly*, 19(2), pp. 139–158. Available at: https://doi.org/10.1080/10573560308222.

Pieters, R. (2017) 'Meaningful mediation analysis: plausible casual inference and informative communication', *Journal of Consumer Research*, 44(3), pp. 692–716.

Pinot de Moria, A., Wheadon, C. and Christodoulou, D. (2022) 'The classification accuracy and consistency of comparative judgement of writing compared to rubric-based teacher assessment', *Research in Education*, 113(1), pp. 25–40. Available at: https://doi.org/10.1177/00345237221118116.

Pollitt, A. (2012) 'The method of Adaptive Comparative Judgement', *Assessment in Education: Principles, Policy & Practice*, 19(3), pp. 281–300. Available at: https://doi.org/10.1080/0969594X.2012.665354.

Purewal, S. (2017) 'Shakespeare in the Classroom: to be or not to be?', *Warwick Journal of Education - Transforming Teaching*, 1, pp. 26–35.

Schunk, D.H. and DiBenedetto, M.K. (2021) 'Chapter Four - Self-efficacy and human motivation', in A.J. Elliot (ed.) *Advances in Motivation Science*. Elsevier (Advances in Motivation Science), pp. 153–179. Available at: https://doi.org/10.1016/bs.adms.2020.10.001.

Singh, A., Uwimpuhwe, G., Vallis, D., Akhter, N., Coolen-Maturi, T., Higgins, S., Einbeck, J., Culliney, M. and Demack, J. (2023) *Improving Power Calculations in Educational Trials*. London: Education Endowment Foundation. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/methodological-research-and-innovations/Work_Package_2023-WP6_18_09_2023_FINAL.pdf?v=1713850351.

Slavin, R.E., Lake, C., Inns, A., Baye, A., Dachet, D. and Haslam, J. (2019) *Writing Approaches in Years 3 to 13: Evidence Review*. London: Education Endowment Foundation. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/documents/guidance/Writing_Approaches_in_Years_3_to_13_Evidence_Review.pdf?v=1733134727.

Smet, M. (2022) 'Professional Development and Teacher Job Satisfaction: Evidence from a Multilevel Model', *Mathematics*, 10(1), p. 51. Available at: https://doi.org/10.3390/math10010051.

Stone, G., Andrade, J., Martin, K. and Styles, B. (2022) *Helping Handwriting SHine - Evaluation Report*. London: Education Endowment Foundation. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/Helping-Handwriting-Shine-Addendum-Report-Final.pdf?v=1743091937.

Torgerson, C.J., Ainsworth, H., Bell, K., Elliott, L., Fountain, I., Gascoine, L., Hewitt, C.E., Kasim, A., Kokotsaki, D. and Torgerson, D.J. (2018) *Calderdale Excellence Partnership: IPEELL - Evaluation report and executive summary*. London: Education Endowment Foundation. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/IPEELL_1.pdf?v=1743596545.

Troia, G.A., Harbaugh, A.G., Shankland, R.K., Wolbers, K.A. and Lawrence, A.M. (2013) 'Relationships between writing motivation, writing activity, and writing performance: effects of grade, sex, and ability', *Reading and Writing*, 26(1), pp. 17–44. Available at: https://doi.org/10.1007/s11145-012-9379-2.

VERBI Software (2021) 'MAXQDA 2022'. Berlin: VERBI Software.

Ward, S.C. and Connolly, R. (2008) 'Let them eat Shakespeare: prescribed authors and the National Curriculum', *The Curriculum Journal*, 19(4), pp. 293–307. Available at: https://doi.org/10.1080/09585170802509880.

Williams, B.W., Kessler, H.A. and Williams, M.V. (2014) 'Relationship Among Practice Change, Motivation, and Self-Efficacy', *Journal of Continuing Education in the Health Professions*, 34, p. S5. Available at: https://doi.org/10.1002/chp.21235.

Williams, D. and Rhodes, R.E. (2016) 'The Confounded Self-Efficacy Construct: Review, Conceptual Analysis, and Recommendations for Future Research', *Health psychology review*, 10(2), pp. 113–128. Available at: https://doi.org/10.1080/17437199.2014.941998.

Zenker, F. and Kyle, K. (2021) 'Investigating minimum text lengths for lexical diversity indices', *Assessing Writing*, 47, p. 100505. Available at: https://doi.org/10.1016/j.asw.2020.100505.

Zumbrunn, S., Marrs, S., Broda, M., Ekholm, E., DeBusk-Lane, M. and Jackson, L. (2019) 'Toward a More Complete Understanding of Writing Enjoyment: A Mixed Methods Study of Elementary Students', *AERA Open*, 5(2), p. 2332858419850792. Available at: https://doi.org/10.1177/2332858419850792.