Writing Roots
Evaluation Protocol

Evaluator (institution): RAND Europe and University of Leeds
Principal investigator(s): Elena Rosa Speciani

Template last updated: December 2022

Evaluation summary

Project title	**Using the Writing Roots programme to improve writing attainment, a two-armed cluster randomised trial**
Developer ***(Institution)***	Literacy Tree
Evaluator ***(Institution)***	RAND Europe and University of Leeds
Principal investigator(s)	Elena Rosa Speciani
Protocol author(s)	Anna Brian, Louise Tracey, Erin Dysart, Paula Clarke, Rebecca Mead, Elena Rosa Speciani
Trial design	Two-arm cluster randomised controlled trial with random allocation at school level
Trial type	Efficacy
Pupil age range and Key Stage	6 – 7 (Year 2, KS1), 9 – 10 (Year 5, KS2)
Number of schools ***(at design stage)***	130
Number of pupils ***(at design stage)***	7,800
Primary outcome measure and source	Writing attainment, Writing Assessment Measure (WAM)
Secondary outcome measure and source	Writing self-efficacy, Writing self-efficacy Measure (WSEM)

Protocol version history

Version	Date	Reason for revision
1.2 **[*latest*]**		
1.1		
1.0 **[*original*]**	02/07/2025	N/A

Contents

Tables

Figures

Background and Objectives

Writing is a complex cognitive skill that requires the integration of linguistic, motor, and executive functions (Berninger & Winn, 2006). Research has consistently demonstrated that targeted interventions can improve writing outcomes (for example, Graham & Perin, 2007a; Graham et al., 2012; Finalayson & McCruddon, 2019), particularly when they are implemented systematically across schools (Graham et al., 2012). Writing interventions often focus on enhancing pupils' ability to plan, draft, revise, and edit their work, which are key components of effective writing (Hayes, 1996; Rogers & Graham, 2008).

Writing attainment has been a persistent concern in UK education, with recent Ofsted reports highlighting the need for innovative approaches to improve literacy outcomes (Ofsted, 2022). The Department for Education (DfE) has emphasised the importance of evidence-based interventions in addressing literacy gaps, particularly in the wake of disruptions caused by the COVID-19 pandemic (DfE, 2021).

The Writing Roots programme is a structured writing intervention developed by Literacy Tree. It adopts a book-based pedagogical approach, which diverges from the traditional genre-based methods commonly used in schools. Genre-based approaches often focus on teaching pupils to replicate specific writing styles, such as narrative or persuasive writing, over extended periods (Myhill et al., 2012). While this method has merits, it can limit exposure to diverse writing conventions and may not fully engage pupils. In contrast, Writing Roots leverages a wider range of texts to encourage pupils to explore multiple conventions and apply them creatively in their writing. This aligns with findings that diversity in classroom texts enhances pupil engagement and motivation (O'Leary et al., 2024). Motivation and enjoyment in writing are strongly correlated with improved writing outcomes, as noted by Ofsted (2022) and Slavin et al., (2019). This diverse approach is particularly relevant given the DfE's commitment to improving cultural capital in schools (DfE, 2019).

The programme also addresses self-efficacy, a critical factor in writing development. Bandura's (1997) Social Cognitive Theory highlights the importance of self-efficacy in academic achievement, suggesting that pupils who believe in their ability to succeed are more likely to engage in challenging tasks and persist in the face of difficulties. Writing Roots aims to foster self-efficacy by providing structured opportunities for success and feedback, which are essential for building confidence in writing.

Despite its apparent potential, Writing Roots has not yet been independently piloted, which underscores the need for rigorous evaluation. And while the evidence base for structured writing interventions is relatively strong, with numerous studies demonstrating their effectiveness in improving writing outcomes (Graham et al., 2012; Slavin et al., 2019), the specific approach adopted by Writing Roots—using a book-based pedagogical framework—has not been extensively studied. While there is evidence supporting the use of diverse texts to enhance engagement (O'Leary et al., 2024), the impact of this approach on writing outcomes remains less clear. Literacy Tree's extensive delivery of the programme to 1168 schools, with 80+ receiving the INSET training package, provides a strong foundation for an efficacy trial. The refinement of the intervention through collaboration with schools and the establishment of clear fidelity criteria further enhance its suitability for evaluation.

The planned evaluation, including impact, implementation, and cost analyses, will provide valuable insights for policymakers and educators. By assessing the programme's efficacy, mechanisms of impact, and cost-effectiveness, the trial will contribute to the evidence base for writing interventions and inform decisions about scaling and adoption.

The evaluation of Writing Roots will assess its efficacy in improving pupils' writing capabilities and self-efficacy compared to business as usual. The trial is a two-arm, cluster-randomised design, with school-level randomisation. Schools assigned to the intervention group will implement Writing Roots for all pupils in Years 1–6, while control schools will continue with business as usual. The impact evaluation (IE) will focus on Year 2 and Year 5 pupils to capture attainment and progression across Key Stages 1 and 2 while minimising the burden on schools and avoiding disruption to Year 6 SATs preparation.

The implementation and process evaluation (IPE) will explore how the intervention is delivered, the mechanisms driving observed outcomes, and how best to implement writing interventions in a whole-school context. This will involve triangulated data collection methods to compare trial arms and capture stakeholder experiences efficiently. Some researchers have raised concerns about the scalability and sustainability of whole-school interventions, particularly in terms of staff training and curriculum integration (Loxley et al., 2007; Harris & Jones, 2017). These challenges highlight the importance of the implementation and process evaluation (IPE) component of the trial, which will examine how the intervention is delivered and identify mechanisms that contribute to its success or limitations. A cost evaluation will assess the programme's value for money, providing insights for schools and policymakers.

Intervention

Name: Writing Roots

Why (theory/ rationale):

The writing curriculum, as conventionally delivered in schools, fails to adequately engage pupils in writing, focusing on one genre or type of text for long periods of time with grammar taught out of context of writing styles. The aim of the Writing Roots programme is for pupils to be able to write in a variety of styles and conventions for diverse audiences and purposes, and to develop a broader knowledge of children's literature.

Research from the Book Trust (O'Leary et al., 2024) suggests that pupil's engagement is improved by the use of diverse and inclusive books in the classroom. This corresponds with anecdotal evidence from schools already participating in Literacy Tree programmes that pupils are engaged by the broad array of texts. This provides the opportunity for pupils to apply different conventions to their own writing, strengthening their ability and adaptability. Previous teacher and school feedback to Literacy Tree has indicated that the Writing Roots resources provide a clear understanding of how to sequence lessons and develop pupils' knowledge in different types of writing, resulting in better ability in writing for different purposes.

An Ofsted (2022) review of English teaching provides evidence for a strong correlation between pupils' motivation to write and their skills as writers, with motivation improved by 'writing for real audiences and purposes' and 'choice of topic', both of which are critical to Writing Roots. An EEF evidence review (Slavin et al., 2019) similarly identifies pupils' motivation to write and enjoyment of writing for self-expression as key characteristics of programmes that successfully promoted good writing outcomes.

A number of elements of the Writing Roots intervention have been identified as being critical to positive writing progression. Programmes which combine writing with reading, as Writing Roots does, have similarly been found to promote a positive writing outcome (for example,

Anders et al., 2021), with some evidence that Writing Root's focus on explicit vocabulary teaching could be predictive of academic attainment (OUP, 2023). Writing Roots' explicit teaching of spelling has been identified in EEF Guidance as important for Key stage 1 and the programme's focus on promoting high-quality dialogue and providing opportunities to practice writing for different audiences has been identified as important for Key stage 2 (EEF, 2020; EEF, 2021).

Who (recipients and provider):

The delivery team, Literacy Tree, will provide the Writing Roots programme to schools, through training and lesson plans. Literacy Tree consultants, who have all previously been teachers, school leaders and Local Authority Advisors in England, will engage all teaching staff delivering the programme. Writing Roots is a whole school programme, so Year 1 to Year 6 pupils and teachers in all schools assigned to the treatment group will be involved in delivering Writing Roots. School leaders are expected to support the implementation and delivery of the programme, for example by considering the visibility of writing in their physical environment and taking steps to embed a culture of writing in their schools.

The delivery team will seek to recruit 130 schools for the purposes of the evaluation. Schools' eligibility to participate in the trial are outlined in the below section on 'Participants'. As we are interested in how the programme works in the context of areas with high socio-economic disadvantage, we are aiming to recruit 50% of the schools in the evaluation sample from Education Investment Areas (EIAs). Half of the settings will be randomised into the treatment group (see 'Randomisation' section for description of stratified randomisation process) and receive the Writing Roots programme in the 2025/2026 academic year. The other half in the control group will have the option of accessing the programme in the 2026/2027 academic year but will not be exposed to the Programme in 2025/2026 so as not to hamper the integrity of the evaluation.

Whilst Writing Roots is a whole-school programme, data for the impact evaluation will be collected only from pupils who will be in Years 2 and Year 5 in the 2025/2026 academic year. These two year groups have been chosen to ensure that attainment and progression is captured across both Key Stages. Other year groups were excluded in order to avoid overburdening schools during data collection, with particular care taken to avoid data collection in year groups where preparation for SAT assessments necessarily must take priority, such as Year 6.

What (materials and procedures):

Literacy Tree consultants will deliver an online half-day launch event for teaching staff in participating schools. This will take the form of an INSET session and is delivered online so that schools across England can attend on a date convenient to them. Catch-up materials will be available online for teaching staff who are not able to attend in person. The aim of the launch event is to introduce teachers in participating schools to Literacy Tree's 'Teach Through a Text' package, build knowledge of different writing conventions, and demonstrate teaching of grammar, and literary language through shared and modelled writing. Year group-specific 'Teach Through a Text' Training will take place in September, January and April of the delivery period. This will be delivered online and focus on the skills relevant to pupils' expected writing level.

To support delivery, schools receive membership to the Literacy Tree Platform, copies of all the texts, curriculum maps, progression documents, and access to an online classroom toolkit. Membership grants practitioners' access to different resources for teaching and assessment, discounts on books, and a book lending service, as well as the professional development they access through Writing Roots. Teachers will deliver the programme through carefully planned lesson sequences allowing pupils to explore texts through interactive activities ('discovery points'), discussion, drama, and debate. These plans will involve different writing styles and audiences with targeted writing outcomes— for example, grammar, punctuation, and sentence structure—chosen for their relevance to particular points in the text, with new writing styles being modelled by teachers. Lesson sequences take a layered learning approach whereby pupils are given opportunities to practice new skills and meet learning objectives multiple times in different contexts, building from shorter to longer writing outcomes. Participating teachers are also able to attend half-termly online planning support sessions with Literacy Tree consultants. Schools can also choose to access additional consultancy at an extra cost.

Where (location)

The schools participating in the trial will be recruited from all over England, including the 55 local authorities that comprise the EIAs.

How (format):

All training (INSET launch event and 'Teach Through a Text' sessions) and planning support sessions will be delivered by Literacy Tree consultants online. The lesson plans and texts will be delivered to pupils by teachers.

When and how much (dosage):

The intervention will take place across the whole school year, with the whole school INSET for the treatment group taking place at the beginning of the 2025/2026 autumn term. Three year-group specific training sessions ('Teach Through a Text') will be held for each of the year groups from Years 1-6: one in September, one in January and one in April. In addition, teachers can book up to six online planning support sessions. Pupils from Year 1 to Year 6 receive one hour of Writing Roots daily.

Tailoring (adaptation):

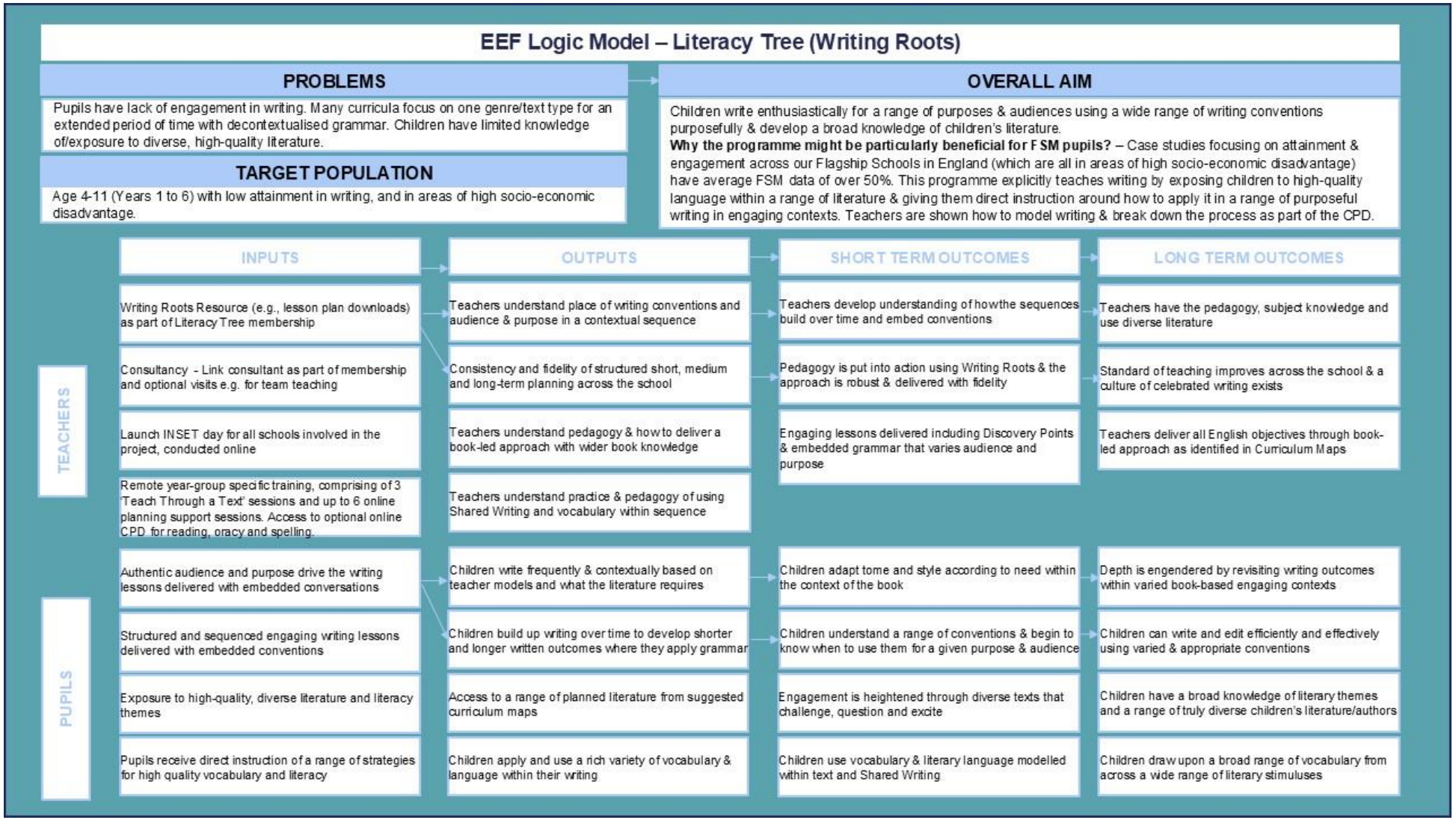
The texts and lesson plans are selected and designed to be appropriate for each year that Writing Roots is delivered to (Years 1 through to 6). Schools have the option of accessing additional online CPD for Reading, Oracy and Spelling, although this is not part of the core programme and is not expected to be included in the delivery by treatment schools. Ways in which teachers adapt their delivery for pupils of different abilities, and with English as an additional language (EAL) and special educational needs and disabilities (SEND), will be evaluated through the IPE.

Control condition:

The control condition will be business-as-usual.

The logic model below (**Error! Reference source not found.**) was developed in collaboration with the delivery team and agreed during the set-up stage.

Figure 1: Logic Model



Methods

Research questions

Primary Research Questions

RQ1. What is the impact of Writing Roots on writing outcomes for pupils in Year 2 as measured by the Writing Assessment Measure (WAM)?
RQ2. What is the impact of Writing Roots on writing outcomes for pupils in Year 5 as measured by the WAM?

Secondary Research Questions:

RQ3. What is the impact of Writing Roots on writing outcomes for pupils in Year 2, as measured by the following WAM subtests:
  a. Vocabulary
  b. Organisation and overall structure
  c. Ideas
RQ4. What is the impact of Writing Roots on writing outcomes for pupils in Year 5, as measured by the following WAM subtests:
  a. Vocabulary
  b. Organisation and overall structure
  c. Ideas
RQ5. What is the impact of Writing Roots on writing self-efficacy for pupils in Year 5, as measured by the Writing Self-efficacy Measure (WSEM)?
RQ6. What is the impact of Writing Roots on writing outcomes for pupils in Year 2 as measured by the WAM on pupils
  a. Who are eligible for FSM
  b. Who are SEND
RQ7. What is the impact of Writing Roots on writing outcomes for pupils in Year 5 as measured by the WAM on pupils
  c. Who are eligible for FSM
  d. Who are SEND
RQ8. Are there spillover effects of Writing Roots on broader pupil attainment over the long term, as measured by pupils' KS2 attainment in reading and maths

Trial design

Table 1: Trial design

Trial design, including number of arms		Two-armed randomised controlled trial
Unit of randomisation		School
Stratification variables (if applicable)		Region/EIA
	Variable	Writing attainment

Primary outcome	**Measure** (instrument, scale, source)	Writing Assessment Measure (WAM), 0-28 (Dunsmuir et al., 2015).
Secondary outcome(s)	**Variable(s)**	Writing attainment on three subtests: 1) vocabulary, 2) organisation and overall structure, and 3) ideas. Writing self-efficacy (for Year 5 only) Reading and Maths Attainment on KS2 SATs (Year 5 cohort only, at 1 year post intervention)
	Measure(s) (instrument, scale, source)	Relevant subsets of the WAM, 0-4 (Dunsmuir et al., 2015). Writing Self-Efficacy Measure (WSEM), 0-100 (Bruning et al., 2013) Scaled KS2 SAT scores on reading and maths
Baseline for primary outcome	**Variable**	Writing capabilities
	Measure (instrument, scale, source)	Writing Assessment Measure (WAM), 0-28 (Dunsmuir et al., 2015).
Baseline for secondary outcome	**Variable**	Three subtests: 1) Vocabulary, 2) organisation and overall structure, and 3) ideas.
	Measure (instrument, scale, source)	Relevant subsets of the WAM, 0-4 (Dunsmuir et al., 2015).

Participants

The study participants will be pupils in Year 2 and Year 5 attending trial schools in the year of the intervention. Parents or carers are provided with a parent information sheets and privacy notices and given the opportunity to opt out of the trial prior to baseline collection or withdraw their child from the evaluation at any time. While the intervention will be delivered at the whole school level, the EEF is keen to understand the effects of the intervention on Key Stages 1 and 2, so we will be conducting the analysis on these two year groups to capture both the immediate and long term effects of Writing Roots.

In Year 2 pupils are developing foundational skills in spelling, grammar, and sentence structure, which are crucial for later writing proficiency (Berninger et al., 2002). In doing so, they are transitioning from learning to write to writing to learn and so more likely to engage with applying conventions, style, and tone to their own writing. In Year 5, pupils are typically developing higher-order writing skills (Graham & Perin, 2007b), such as organisation and argumentation and are in the preparatory phase for high-stakes formal assessment in Year 6.

Schools will be recruited from England, provided they meet the following criteria:

- Are a state-funded mainstream primary school with all year groups from Year 1 to Year 6 (whole primary school), with no more than two year groups in mixed-year classes.
- Not participating in the following EEF trials: Pathways Literacy, Power of Reading, The OTTO Club, Rehearsal Room Writing, or any whole-school trials.
- Are not currently using Literacy Tree's Writing Roots programme and have not participated in any Literacy Tree programme or training for the past two years.

The EEF is keen to include schools from EIAs, which span 55 Local Authorities (LAs). Recruiting 50% from EIAs will enhance understanding of Writing Roots' impact both within and outside these areas. We have used this approach successfully on the Lexia evaluation (Brown et al., 2022). Randomisation will be stratified by region that will account for EIA, ensuring that an equal number of schools in each EIA will be assigned to the treatment and control groups.

There will also be eligibility requirements for pupils to be included in the analytical sample. All pupils in intervention schools will receive the intervention but children with Education, Health, and Care Plan (EHCP) will not be included in data collection or analysis. While the Writing Roots programme is accessible for all, it is based on age related expectations, so under efficacy trial conditions, pupils with EHCP will not be included in the data collection or analysis. This will still enable SEND subgroup analysis as not all SEND pupils have an EHCP; however, the SEND subgroup analysis will not be representative of the SEND population within the participating schools and thus not generalisable to all SEND students.

Planned treatment units and how they will be recruited

There will be approximately 65 treatment units (schools) recruited by the delivery team, along with 65 control units. Schools will be recruited to the trial through expressions of interest (EOIs) submitted through Literacy Tree's website, where they can access a school information sheet, outlining the purpose, cost, and eligibility criteria of the trial. Participating schools will then be selected from these EOIs, with the delivery team responsible for screening schools based on the exclusion criteria.

Treatment units will receive the Writing Roots programme worth £3000 at the subsidised cost of £399.50. Schools that are assigned to the control group in the 2025/2026 academic year will have the option of subsidised access to the Writing Roots programme in the 2026/2027 academic year and be given £500 at the end of the evaluation for participation in the data collection (i.e., endline and baseline data surveys and writing samples).

Outcomes

Primary Outcome

The Writing Assessment Measure (WAM)

The primary outcome of the study will be writing attainment. We will measure this using the Writing Assessment Measure (WAM), which will be administered to pupils in their classes. Pupils are presented with a prompt designed to elicit a written narrative response and asked to write for 15 minutes. The version proposed in Dunsmuir et al., (2014) and used in previous

EEF trials marks seven elements of writing on a four-point scale, giving each an equal weight: handwriting, spelling, punctuation, sentence structure and grammar, vocabulary, organisation and overall structure, and ideas. Its use in previous EEF trials (Anders et al., 2021; Stone, 2022) has shown appropriate validity and reliability, as well as being relatively easy to administer and acceptable to schools.

The WAM is a generic measure of writing and while it has been used from 2014 and been marked consistently across pupils, the WAM may not accurately reflect the modern curriculum or outcomes targeted by the Writing Roots programme (e.g. pupil's progression in spelling and handwriting). Concerns have also been raised by the delivery and evaluation team about how engaging the prompt is for pupils, as an unengaging prompt could result in lower quality scripts that do not reflect pupil's true writing ability. As a result, before the main trial, we will conduct a short pilot of a new prompt developed in partnership between Literacy Tree and the evaluation team (see Piloting section below for further details).

In summer 2025, for baseline testing, the WAM will be administered to pupils who will be in Year 2 and Year 5 in the 2025/2026 academic year. This means that pupils in Year 1 and Year 4 will be asked to respond to the prompt. Class teachers will be asked to administer the prompt. This is prior to randomisation so limits the bias that could potentially be introduced. Teachers will return scripts to the evaluation team for marking.

At endline, independent test administrators[1] trained in data collection, data protection, and safeguarding and managed by the University of Leeds will visit schools during Summer 2026 to administer the WAM to ensure there is no teacher influence that would bias the final writing scripts.

Piloting

The new prompt will be piloted in three to five schools recruited by the delivery team and administered to pupils in Year 2 and Year 5 in the 2024/2025 academic year.[2] Pilot schools should not have had significant prior exposure to Writing Roots or Literacy Tree so as to not bias results, and should reflect the intended population to be recruited into the main trial. For example, the schools should not have previously paid for Writing Roots or engaged with other Literacy Tree interventions. The sample, which should contain about 180 scripts will ensure that there are equal numbers of scripts from Year 2 and Year 5 so that the results of the pilot are robust in terms of different developmental stages.

The first aim of this pilot is to ensure that the prompt produces valid scores. Researchers will score scripts according to the original WAM rubric and produce histograms of the descriptive statistics of the scores. The range of outcomes should be similar to what one would expect from educational assessments, which we will evaluate based on the following success criteria:

1. Distribution: The histogram of scores should approximate a normal distribution:
    a. Acceptable skewness values between -1 and +1 (Bulmer, 1979).

[1] Administrators will be drawn from a pool of former teachers and post-graduates with relevant topic experience (i.e., psychology, speech and language therapists)

[2] This means that, while we expect some or all of these pilot schools to also be participating in the main trial, none of the pupils responding to the prompt as part of the pilot will be participating in the main trial.

b. kurtosis values between -1 and +1 (DeCarlo, 1997).

2. Consistency between prompts: The descriptive statistics should be broadly in line with those reported in previous studies, including:

    a. A standard deviation within 10% of those reported in similar studies (Cohen, 1988).

    b. The mean should be within one standard deviation of the mean reported in comparable studies, indicating that the prompt is neither too easy nor too difficult.

We will also conduct this analysis on WAM scores with handwriting and spelling elements removed, which are less crucial outcomes for measurement of the efficacy of the intervention. If the distribution of outcomes from these prompts is valid without these elements, they will be removed so the primary outcome better reflects the intended outcomes of the intervention.

Testing of the prompt will take place before baseline so that the prompt (if valid) can be used at baseline. If the adapted prompt does not satisfy the success criteria it will not be considered appropriate and we will use the original WAM prompt.

Marking

We will seek to use an Artificial Intelligence (AI) marking tool to improve the efficiency and relieve the resource-intensity of marking scripts using the WAM rubric. Advances in natural language processing have improved AI's accuracy in assessing writing, offering detailed data (Shermis & Burstein, 2013) (e.g., raw scores, subtests) compared to methods like comparative judgment (CJ). Using AI also ensures uniform scoring, reducing human bias and enhancing reliability across groups (Williamson et al., 2006).

Preliminary testing as part of the proposal found that feeding individual scripts to RAND Chat, the RAND in-house AI chat bot, produced accurate writing assessment and reasoned justification of scoring. However, it is still vital to verify that AI can manage consistent scoring and produce similar results to human markers on a larger scale, therefore, we will run a piloting process alongside the set-up phase to refine this tool.

We will conduct this pilot using the same scripts as the prompt pilot to streamline the timelines needed for both pilots. The intended sample of 180 scripts will produce sufficient statistical power to detect meaningful patterns and differences, aligning with common practices in pilot studies (Creswell & Creswell, 2017). A sample size of around 100 per group is generally adequate for conducting a Kolmogorov-Smirnov test, which assesses the goodness of fit between the observed and expected distributions, providing insights into the AI's performance consistency. This approach is consistent with precedents in educational research, where similar sample sizes are used to validate tools and interventions (Spandel, 2006). This sample will also be diverse in terms of the developmental stage and writing abilities of the respondents, which is crucial for the testing of the AI tool's adaptability (Luckin et al., 2016).

Scripts from the prompt pilot will be marked using human markers and AI to allow for direct comparison of the reliability of the AI tool's output, through inter-rater reliability metrics. The AI tool will be developed iteratively; starting with ten scripts we will design a prompt that ensures we get relatively high success criteria while factoring in the WAM marking rubric. We will then check the prompt with another 20 scripts, checking that the success criteria are maintained. If

the prompt is not reliable at the larger sample, we will reiterate on the prompt again. Once we get the 20 scripts close to the success criteria, we will check the prompt with another 50 scripts, then 100, until we get to the final sample.

We will use the following validation criteria to ensure that the AI tool is not introducing bias and producing reliable results:

1. Accuracy: AI scores should be within ±10% of human markers' scores in at least 90% of cases. In addition, we will assess whether there are significant differences between the distribution of scores from AI and human markers using a Kolmogorov-Smirnov test.

2. Consistency: Variance in AI scores for the same writing sample should be less than 5%.

3. Reliability: Inter-rater reliability (e.g., Cohen's kappa) should be above 0.75 to be considered good agreement (Landis & Koch, 1977).

4. Efficiency: The AI should mark each writing sample in under 2 minutes.

If the AI tool produces acceptable outputs, RAND will mark the scripts using the AI tool with interrater reliability checks conducted on a sample to confirm scoring and produce further reliability estimates.

Secondary Outcome

Given Writing Roots' emphasis on vocabulary and literacy language modelled on text and shared writing, we are looking at the impact on separate subsets of the WAM, specifically vocabulary, organisation and overall structure, and idea. Looking at these separately allows a better understanding of the extent to which Writing Roots influences some of the fundamental short-term outcomes identified in the logic model in Figure 1.

A significant body of evidence highlights the importance of self-efficacy for successful writing performance (for example, Golparvar & Khafi, 2021). This aligns with broader evidence that the best predictors of academic performance are self-efficacy beliefs in the same academic domains (Gutman and Schoon, 2013). As such, we are using writing self-efficacy as a secondary measure using the Writing Self-Efficacy Measure (WSEM) (Bruning et al., 2013). An adapted version of the WSEM has been used in previous KS1 and KS2 EEF trials and showed a small and positive effect in the Learning about Culture evaluations (Anders et al., 2021). In this trial the distribution of the WSEM had a negative skew at baseline, potentially reducing the size of the estimated effect size.

The WSEM involves 16 statements capturing pupils' perceptions of their writing capabilities, including concepts found on the Writing Roots logic model such as vocabulary use ('I can think of the words I need to write down my ideas') and engagement ('I can keep on writing even when it gets difficult'), with pupils self-reporting on a five-point Likert scale.

We propose to only use the WSEM measure in the Year 5 group. Like the WAM, this test will be administered by independent test administrators. The language of WSEM would need to be adapted for a younger age group if we were to extend the secondary analysis to Year 2, which would require a larger budget and time commitment for piloting. As there is no reliable self-efficacy measure for Year 2, either generic or writing-specific, we propose to restrict this

measure to Year 5. We will explore Year 2 feelings about writing as part of the IPE using questions drawn from the WSEM.

Longitudinal outcomes

The longitudinal analysis is focussed on understanding whether there are unintended consequences (positive, or negative) in pupil attainment as a result of taking part in Writing Roots. This is based on the fact that some studies have found that interventions targeting one domain can sometimes lead to unintended trade-offs in others, particularly if instructional time is reallocated disproportionately (Van der Kleij et al., 2015). However, there is also evidence that interventions targeting literacy often improve outcomes in other areas, such as maths, due to enhanced reasoning and communication skills (Slavin et al., 2009).

Maths and Reading outcomes will be measured using scaled KS2 SAT scores on reading and mathematics from the NPD, for the Year 5 cohort only, at one year post intervention endline (i.e., when the Year 5 cohort are finishing Year 6).

Baseline Measures

We will use the WAM at baseline, administered by teachers and marked by the AI tool, for several reasons: We know from previous trials that the WAM has a relatively low administration burden, with teachers in Year 2 and Year 5 successfully administering the WAM at baseline (Stone et al., 2022). Using the same test at baseline and endline will ensure higher pre-test/post-test correlations which improves power within a relatively limited sample size. Efficiencies could be made by using administrative data from the National Pupil Database (NPD), but this would be problematic for the Year 5 cohort as they are missing Early Years Foundation Stage Profile (EYFSP) data owing to being cancelled because of the COVID-19 pandemic, and KS1 data is not granular enough to use for baseline. We know previous trials (Lord et al., 2018) have tried to collect granular KS1 data directly from schools but have not been successful.

Sample size

We note that WAM pre-test/post-test correlations at Level 1 are 0.82 or $r^2 = 0.67$ based on test-retest correlations with a one-month test gap (Dunsmuir et al., 2015). Given that we propose to use the same test at baseline and endline (a year apart) there is a high chance that correlations will be higher than those seen in other EEF trials of KS2 English which are around 0.6 ($r^2 = 0.36$) (Singh, 2023). However, we conservatively assume correlations of 0.7 ($r^2 = 0.49$). We also note that in previous EEF KS2 English trials there is a strong relationship between pre-test/post-test correlations at the pupil and school level (Level 1 = 0.60, Level 2 = 0.57) (Singh et. al, 2023) and, as such, we have assumed that at Level 2 the r^2 is 0.45. We use this across Year 5 and Year 2 calculations and for FSM pupils as well. While we acknowledge there will likely be differences across ages and subsets of pupils, we do not have enough data to generate empirical assumptions.

Singh et al., (2023) recommend an ICC of 0.10. Our own review of EEF trials with writing outcomes, produces an average ICC for Year 5 of 0.17 across all and FSM pupils; for Year 2 it is 0.12 for all and 0.1 for FSM pupils. To be conservative, for our sample size calculations we use 0.13 for all Year 5 pupils, 0.12 for all Year 2 pupils, and 0.1 for both Year 2 and Year

5 FSM pupils. We also assume a sample size of 30 pupils in each class, and a sample of 8 pupils eligible for FSM in the last 6 years.[3]

Using these parameters the MDES for all pupils and FSM pupils are presented below. All MDES calculations were made using PowerUp!.

Table 2: Year 2 MDES/sample size calculations

		Overall	FSM
Minimum Detectable Effect Size (MDES)		0.141	0.164
Pre-test/ post-test correlations	level 1 (pupil)	0.70	0.70
	level 2 (school)	0.67	0.67
Intracluster correlations (ICCs)	level 2 (school)	0.12	0.10
Alpha		0.05	
Power		0.8	
One-sided or two-sided?		Two-sided	
Average cluster size		30	8
Number of settings	Intervention	65	65
	Control	65	65
	Total	130	130
Number of pupils	Intervention	1950	585
	Control	1950	585
	Total	3900	1170

Table 3: Year 5 MDES/sample size calculations

		Overall	FSM
Minimum Detectable Effect Size (MDES)		0.146	0.164
Pre-test/ post-test correlations	level 1 (pupil)	0.70	0.70
	level 2 (setting)	0.67	0.67
Intracluster correlations (ICCs)	level 2 (setting)	0.13	0.10
Alpha		0.05	
Power		0.8	
One-sided or two-sided?		Two-sided	

[3] Based on the EVER6 FSM eligibility indicator and indications that this is 27.7% of the population (see Campbel et al., 2025).

Average cluster size		30	8
Number of settings	Intervention	65	65
	Control	65	65
	Total	130	130
Number of pupils	Intervention	1950	585
	Control	1950	585
	Total	3900	1170

We also calculated what MDES would be after attrition. Our attrition assumptions come from previous EEF trials of English in primary schools, with attrition of 12% at the school level and 22% at the pupil level, these are presented in the table below.

Table 4: MDES after attrition

	Pupil attrition only	**School attrition only**	**Pupil and school attrition**
Year 2	0.143	0.160	0.163
Year 5	0.147	0.161	0.167

The trial is powered to detect an MDES on the overall sample of between 0.14 and 0.15, well below the 0.2 threshold to attain a maximum 5-padlock security rating. However, the initial MDES calculations presented in Table 2 and 3 do not take into account attrition. Allowing for both pupil and school attrition, MDES could increase to between 0.15 and 0.16. The subgroup analysis on pupils in receipt of free school meals is higher, at 0.164.

Randomisation

Randomisation will be stratified by region, with settings the unit of randomisation and pupils the unit of analysis. This randomisation approach is driven by the whole-school approach of the intervention, which requires that all pupils in each school are assigned to the same condition. Stratification by region will capture whether the setting is part of an EIA, ensuring schools in EIAs are evenly assigned to the treatment and control groups. This will control for some of the variation between settings, improving the precision of the estimate of treatment effect.

The sample will be evenly split between the treatment and control groups. Settings assigned to the treatment group will receive Writing Roots for a subsidised price and be expected to deliver it in the 2025/26 academic year. Settings in the control group will be expected to continue with business as usual and receive a £500 incentive payment after the endline stage, with the option of purchasing the subsidised Writing Roots programme in the 2026/27 academic year.

Randomisation will take place after baseline data collection and was originally planned to be conducted in one ‘batch’ at the end of June 2025 to allow the delivery team to prepare for training in September 2025. However, recruitment issues meant that the recruitment testing window was extended. So as to not confuse schools that had been expecting to be told about

allocation, it was decided by the evaluation team, the EEF, and the delivery team to proceed with block randomisation. There will be two blocks: one randomised at the end of June and another at the end of the school term in July. The size of the two blocks are likely to be different, determined by the number of schools who have i) signed MoUs and DSAs and ii) returned pupil data (names, DoBs, UPNs) and baseline tests by the end of June. All schools who meet this criteria at a specified cut-off date at the end of June will be randomised in block one. All other schools will be randomised in block two.

We will employ stratified randomisation for each block, stratifying on EIA status (i.e., whether a school is in an EIA) to ensure balance of EIA schools. We will check the allocation balance across treatment groups within each block, adjusting the second batch's block size if necessary to maintain overall balance (Schultz et al., 2002). We will employ statistical software (e.g., Stata,) to randomise assignment, to minimise bias and ensures true randomness.

Randomisation in both blocks will be conducted by a RAND researcher who will be blind to treatment allocation. The code used to randomise settings as well as all relevant variables will be recorded and communicated to the implementation team in a PDF file to prevent editing. For transparency we will include details of the block size, stratification variables, and randomisation software in the trial SAP.

Statistical analysis

Primary analysis

The primary analysis will be conducted on an Intention To Treat (ITT) basis. Settings will be analysed as randomised, with all randomised settings and tested pupils being included in the analysis according to their assigned treatment regardless of the treatment actually received, or any deviations in the delivery of the intervention. The ITT approach is inherently conservative as it captures the averaged effect of offering the intervention, regardless of whether the participants complied with assignment. This principle is key in ensuring an unbiased analysis of intervention effects and is in line with the EEF's guidance (EEF, 2022).

The primary outcome will be the WAM total scores (see the 'Outcome measures' section for more detail). The impact of the Writing Roots programme on this primary outcome will be estimated using a two-level multilevel model (pupils and settings) to account for clustering of data with setting region (the stratification variable used in randomisation) and baseline WAM scores controlling for prior writing abilities, as follows:

$$\mathrm{Y_{ij}} = \beta_0 + \tau\,\mathrm{WR_j} + \beta_1\,\mathrm{Z_j} + \beta_2\,\mathrm{X_{ij}} + \mathrm{u_j} + \mathrm{e_{ij}}\ (1)$$

Where Y_{ij} represents the WAM total score at endline for pupil i in school j. The main outcome of interest is τ which represents the estimate of the effect a school being assigned writing roots, represented by WR_j, a binary indicator (equalling 0 if the school is assigned to the control group and 1 if it assigned to the treatment group) has on the primary outcome. β_0 is the cluster-level coefficient for the slope of a predictor on writing skills; β_1 is the impact of the setting-specific characteristics, specifically region, captured in $\mathrm{Z_j}$; β_2 is the impact of pupil-level characteristics, specifically baseline scores, captured in $\mathrm{X_{ij}}$. Setting level residuals are captured in $\mathrm{u_j}$ and individual-level residuals are captured in $\mathrm{e_{ij}}$.

Equation (1) is known as a 'random intercepts' model because $\beta_{0j}=\beta_0+u_j$, the setting-specific intercept for school j, is random (it is a number that can take any value) in nature, with an assumed distribution $\beta_{0j} \sim_{i.i.d} N(\beta_0, \sigma_u 2)$. The effect size (Hedge's *g*) will be standardised using unconditional variance in the denominator and confidence intervals will be reported to communicate statistical uncertainty in line with EEF guidance. This will tell us the average effect of the intervention on pupil's writing outcomes in treatment settings compared to those in control settings.

This analysis will be conducted separately for Year 2 and Year 5 pupils. KS1 and KS2 pupils are at different development stages, impacting their learning and response to interventions, necessitating separate analyses (Goswami, 2015; EEF, 2020). Separate analyses will also improve statistical precision and power by reducing variability. Distinct analyses allow for clearer interpretation of results, making it easier to identify specific effects for each year group.

Secondary analysis

The following secondary analyses are planned:

(1) An analysis of vocabulary, organisation and overall structure, and ideas, using subsets of the WAM marking criterion

(2) an analysis of writing self-efficacy using the WSEM

Both outcomes are described in the 'Outcomes' section above.

The secondary analysis using WAM subsets will be assessed using the same specification in equation (1) outlined in the 'Primary analysis section'. However, we will substitute the endline measure of the WAM score for vocabulary, organisation and overall structure, or ideas for the outcome $\mathrm{Y_{ij}}$ and the respective baseline score for $\mathrm{X_{ij}}$.

We will be conducting two secondary analyses using the WSEM as the outcome of interest. The first will be similar to the primary analysis specification outlined in equation (1) using WSEM as the endline outcome ($\mathrm{Y_{ij}}$) and keeping WAM as the baseline measure, accounting for pupil-level differences in ability ($\mathrm{X_{ij}}$).

This will be accompanied by a sensitivity analysis where the effect of Writing Roots on self-efficacy will be evaluated using an endline only model, using the following specification.

$$Y_{ij} = \beta_0 + \tau WR_j + \beta_1 Z_j + u_j + e_{ij} \quad (2)$$

In this specification, Y_{ij} is the WSEM outcome at endline and pupil-level characteristics are not controlled for as they are in equation (1); however, other coefficients remain the same, with the coefficient τ on the Writing Roots dummy (WR_j), still signifying the main outcome of interest: the estimated impact of the Writing Roots programme on writing self-efficacy.

Exploratory analysis

We will look at the impact of Writing Roots at the school level by combining data from Year 2 and Year 5 and will carefully consider how to manage split-cohort data, accounting for confounding variables (Ray et al., 2022) either by pooling (i.e., merging data from both cohorts) or analysing cohorts separately and aggregating effect sizes through meta-analysis.

Sub-group analysis

A sub-group analysis on pupils from disadvantaged backgrounds, as measured by eligibility for FSM, is critical, given the existing focus of the EFF on this sub-group. However, given the trial has been powered to detect a moderate effect on the overall sample, there is a risk that the FSM sub-group analysis may be under-powered due to the smaller sample sizes. We will collect data from the NPD on pupils' eligibility for FSM using the *EVERFSM_6_P* variable from the NPD, in line with EEF guidance. Analysis will be undertaken with the binary *EVERFSM_6_P* variable used as a moderator, therefore using the whole trial sample.

While we are excluding pupils with EHCP from the analysis, pupils who are SEND without EHCP will still be included, enabling a subgroup analysis of the impact of Writing Roots on SEND pupils, who will be identified using NPD data (using variable SENprovision). We will also conduct subgroup analysis on the impact on EAL pupils (using variable LanguageGroupMajor). However, it should be noted that the EAL label captures pupils with highly variable levels of English.

Analysis of the subgroups will use two approaches, as suggested in EEF analysis guidance (EEF, 2022). The first will run the primary model given in Equation (1) on the FSM, SEND, and EAL subgroups only. Effect sizes and statistical uncertainty will be calculated on these subgroups following the procedure outline in the above section on 'Primary analysis'.

The second approach will make use of the entire sample. The treatment effect on each subgroup will also be estimated using an interaction model, here using FSM as an example:

$$Y_{ij} = \beta_0 + \tau WR_j + \beta_1 FSM_i + \beta_2 (FSM_i * WR_j) + \beta_3 Z_j + \beta_4 X_{ij} + u_j + e_{ij} \quad (3)$$

This is the same model specification as in equation (1), with the addition of the FSM_i indicator of disadvantage and an interaction term combining FSM eligibility and treatment allocation ($FSM_i * WR_j$). The primary coefficient of interest in the interaction model is β_2, which can be interpreted as the additional treatment effect experienced by pupils from disadvantage background: a positive β_2 is indicative of a treatment acting as a 'gap-closer' and a negative β_2 indicative of treatment acting as a 'gap-widener'. The treatment effect size will be calculated by hand using the coefficients in the interaction models and the unconditional standard deviation of the FSM sub-sample[4], according to EEF guidance (2022), and compared with that calculated from the model on the FSM sub-sample. We will repeat this analysis for the SEND and EAL subgroups.

Given previous evaluations have noted marked differences between low and high attainers (Slavin et al., 2019), we will also look at whether there are differential impacts for learners from different attainment quartiles at baseline.

Analysis in the presence of non-compliance

As the ITT approach is inherently conservative, capturing the averaged effect of offering the intervention, we also propose to look at treatment effects in the presence of compliance. This additional analysis measures the average effect of fully compliant participation in Writing Roots on literacy outcomes. The definition of compliance and how it will be measured has been

[4] This is calculated according to the following formula: $\frac{\tau WR_j + \beta_2 (WR_j * FSM_i)}{sd}$

agreed with the delivery team and EEF. Compliance will be a composite measure formed on various data points that will come from Literacy Tree:

Metric	Source of data	Threshold for compliance
Lesson plan downloads	Number of downloads by a specific user captured by the Literacy Tree website (backend data source) provided by Literacy Tree	At least 9 out of 12 Writing Roots lesson plans downloaded for Year 2 and at least 9 out of 12 Writing Roots lesson plans downloaded for Year 5.
Attendance at INSET	Training attendance logs, or (where catch-up is needed) viewing statistics of individual training catch up videos from Literacy Tree website. Both provided by Literacy Tree.	All teachers attend the INSET day or watch at least 70% of an INSET catch-up video
Attendance at 'Teach Through a Text' sessions	Attendance logs from Literacy Tree.	All teachers attend two out of three 'Teach Through a Text' sessions

This will be a binary measure whereby settings will be marked as compliant if all of the above criteria are met and non-compliant otherwise. In a situation of imperfect compliance — not all intervention settings are deemed compliant according to the above criteria — we will undertake a complier average causal effect (CACE) analysis, using two-stage least squares (2SLS) estimation to recover the local average treatment effect (LATE) of attending a compliant setting on writing outcomes.

Additional analyses and robustness checks

Missing data

Missing data can arise from item non-response or attrition of participants at setting and pupil levels. Even though it is important to include all data, it can be problematic to apply the ITT principle if we are not able to complete follow-up testing for all randomised settings or pupils. To better understand the pattern of missing data and its impact on the analysis, we will explore the extent of missingness, and whether there is a pattern in missingness.

Attrition across both trial arms will be explored as a basic step to assess bias. For less than 5% missingness overall, we propose to only carry out a complete-case analysis, regardless of

the missingness mechanism. To gauge systematic differences in missingness, we propose to model missingness at follow-up as a function of baseline covariates, including treatment status. This will allow us to further investigate the pattern in missingness. Depending on the pattern of missingness, multiple imputation may be implemented according to EEF guidance (2022). However, should data be missing not at random, multiple imputation will not be sufficient to generate unbiased estimates of the treatment effect, and sensitivity analysis will be carried out and reported alongside the headline impact estimates.

Longitudinal analysis

There is potential for and spillover effects on participants on other learning outcomes (both positive and negative). For example, deep engagement with texts as part of the programme could improve reading skills. We will use data in the NPD to capture KS2 reading and mathematics scaled scores for the Year 5 (when they are in Year 6) to look at potential unintended outcomes. Specifically, we will request the KS2_READSCORE for reading and the KS2_MATSCORE for maths.

Data from the NPD will be accessed through the Integrated Data Service (IDS) by members of the evaluation team who are ONS' accredited researchers. Using pupil level data (name, date of birth, gender, unique pupil number), data on the original Year 5 pupils' KS2 attainment at the end of Year 6, as measured by scaled scores on reading and maths SATs, will be linked to their WAM and WSEM scores (a similar process to that for subgroup analysis). The evaluation team will analyse the long term outcomes of pupils who are in Year 5 in the delivery year in line with Model (1), with KS2 attainment as the outcome and WAM baseline scores to control for child ability.

Implementation and process evaluation (IPE) design

Research questions

The IPE takes place over two years. The research questions for the IPE in the main trial year (2025-2026) (IPERQs1-6) have been designed to understand how implementation of the programme occurs throughout the trial, with a focus on the assumed casual mechanisms within the programme's Theory of Change, and how this may influence impact, as addressed in the IE for the main trial. The focus in the follow-on study (2026-2027) is on the extent to which the programme is embedded within schools following the intervention year and to explain the findings of the longitudinal strand of the IE (i.e. the outcomes from the Year 6 Standard Assessment Tests (SATs) in the year following the main trial). The research questions for the follow-on study (IPERQs7-9) aim to provide longer term answers to some of the key questions relating to implementation asked in the trial year (in particular, IPERQs 2-4). The research questions (alongside the associated implementation dimensions)[5] are as follows:

Main trial

IPERQ1. To what extent do schools engage with the training? *(Fidelity, dosage, responsiveness)*

[5] See Humphrey et al. (2019), p.6 for definitions of the implementation dimensions.

a. Are schools able to release teachers for training and do all teachers engage with the training?

b. To what extent does the quality of training accessed vary (e.g. whole school INSET, 'Teach Through a Text' training sessions)?

c. What are the barriers and facilitators of attendance at training?

d. What levels of support do schools feel they need to implement Writing Roots and do they receive it (e.g. additional consultancy)?

IPERQ2. To what extent do teachers and schools implement Writing Roots as intended? *(Fidelity, quality, reach, adaptations)*

a. How far is the programme adopted as a whole school approach (including SLT and Literacy Lead support) and how does that impact on school culture (with particular regard to reading and writing culture)?

b. What variability exists in implementation in terms of the key programme components (e.g. lesson plans provided, use of recommended texts, explicit teaching of spelling, embedded explicit vocabulary teaching)?

c. How does variability differ by year group and/or pupil factors (including FSM, EAL)? Do teachers differentiate their delivery based on pupil prior attainment and do they utilise the mixed year group planning resources in mixed year group classes?

d. What are the barriers and facilitators of implementation at the school and class level?

IPERQ3. To what extent does the programme and associated training impact on teachers' knowledge and confidence in using a book-led approach to teach writing conventions across different literacy themes? To what extent does this inform their pedagogical approach? *(Quality, Responsiveness)*

IPERQ4. What impact does the programme have on pupils' engagement with texts, and their confidence and motivation to write? *(Reach, responsiveness)*

a. Is the selection of texts perceived to be appropriate for pupils (with particular regard to representing diverse pupil backgrounds and experiences)? Do pupils develop a broad knowledge of children's literature?

b. Does Writing Roots impact pupils' reading and writing for pleasure?

c. Does Writing Roots increase pupils' stamina in writing (i.e. do they write for longer over time)?

IPERQ5. What is the nature of teaching 'as usual' and to what extent does this differ from the implementation of Writing Roots, including the teaching of spelling, grammar, and punctuation? *(Monitoring of the control condition)*

IPERQ6. To what extent does Writing Roots result in positive or negative unintended consequences for schools, teachers and pupils in other subject domains, particularly in literacy (e.g. Phonics)?

Follow-on study

IPERQ7. To what extent do teachers and schools continue to implement Writing Roots as intended? *(Fidelity, quality, reach)*

a. How far is the programme adopted as a whole school approach (including SLT and Literacy Lead support) and how has this changed during the follow-up year?

b. What variability exists in implementation in terms of the key programme components (e.g. lesson plans provided, use of recommended texts, explicit teaching of spelling, embedded explicit vocabulary teaching)?

c. How does variability differ by year group and/or pupil factors (including FSM, EAL)? Do teachers differentiate their delivery based on pupil prior attainment and do they utilise the mixed year group planning resources in mixed year group classes?

d, What are the barriers and facilitators of continued implementation at the school and class level?

e. Have schools implemented any new writing practices or programmes in the academic year 2026/2027?

IPERQ8. To what extent does the programme and associated training continue to impact on teachers' knowledge and confidence in using a book-led approach to teach writing conventions across different literacy themes? To what extent does Writing Roots continue to inform their pedagogical approach? *(Quality, Responsiveness)*

a. Do teachers' feel the need for any additional support to facilitate continued implementation?

IPERQ9. Is continued implementation and embedding of the programme perceived to have cumulative impacts on pupils' engagement with texts, and their confidence and motivation to write? *(Reach, responsiveness)*

a. Does Writing Roots impact pupils' reading and writing for pleasure?

b. Does Writing Roots increase pupils' stamina in writing (i.e. do they write for longer over time)?

Research methods

During the main trial the IPE will take a mixed-methods approach allowing for triangulation, comparison between arms of the trial, and is designed to capture the experience of key stakeholders. The activities included within the IPE in the trial year are outlined below.

Training observations. To understand Writing Roots more fully and the training associated with the intervention, researchers will attend two whole-school INSET launch events and six year group specific online 'Teach Through a Text' training sessions. INSET events will take place online. Ideally, the two observations of INSET events will involve schools with different demographics (for example, a large primary and a smaller primary school) depending on the final sample. The six observed online 'Teach Through a Text' training sessions will comprise of:

- two training sessions for Year 2 teachers (one in September 2025 and one in April, 2026);
- two training sessions for Year 5 teachers (one in September, 2025 and one in April, 2026);
- one training session for Year 1 teachers (in January, 2026); and
- one training session for Year 6 teachers (in January, 2026).

Where possible, the sessions attended will be randomly selected.

Training observation schedules will be established based on the programme materials and the researchers' understanding of the programme, as established in the Intervention Delivery and Evaluation Analysis (IDEA) meeting during the set-up phase of the project. The observations will include an observer checklist (based around programme training materials) as well as observer notes. These observations will contribute to answering IPERQ1 (a and b), relating to teacher engagement with the training and the extent to which the quality of the training may vary. In addition, researcher attendance at training will contribute to understanding the teacher inputs and outputs relating to the CPD in the Theory of Change. Observations will also assist researchers in answering IPERQ2, particularly implementation of key programme components and potential variability due to year group and/or pupil factors. Although Year 1 and Year 6 are outside of the scope of the IE, including these two year groups is expected to provide the researchers with an understanding of the programme as a whole-school intervention. Year 1 and Year 6 were specifically chosen because both are the focus of national assessments (Year 1 phonics screening check and KS2 SaTs). At least one INSET event and one year-group specific 'Teach Through a Text' training session will be attended by two researchers to ensure inter-observer reliability.

Interviews with trainers. The research team will conduct one-to-one interviews with three Literacy Tree trainers at the end of the delivery period (Summer 2026). The interviews will be designed to understand school and teacher engagement with the training (INSET events and year-group specific 'Teach Through a Text' training sessions) more thoroughly as well as the programme principles. In particular, they will assist in answering IPERQ1 (engagement with training and additional support), IPERQ2 (adaptations and variability between different year groups and according to other pupil factors) and IPERQ3 (impact of the training on teachers' knowledge and confidence in using a book-led approach to teach writing conventions across different literacy themes and their pedagogical approach). Interviews will also enable the researchers to explore more fully the teacher inputs and outputs relating to the CPD in the Theory of Change. The interviews will take place online and the trainers will be sampled based on availability. They will be reassured prior to interview that their responses will remain anonymous.

Routinely collected programme data. Data relating to teacher attendance at online training, in particular the 'Teach Through a Text' training sessions and records relating to teachers' downloading of resources, all collected via Literacy Tree's membership website, will be downloaded by Literacy Tree. Additional data relating to attendance at the INSET events and additional support accessed by schools will also be collected by the delivery team. This data will include which staff attended sessions (INSET event and year-group specific 'Teach Through a Text' training sessions) and the extent to which they engaged with the online resources. It will be used to understand the extent to which the teachers engage with the programme (and the extent to which Writing Roots is adopted as a whole school intervention) and whether or not additional consultancy was accessed by intervention schools (IPERQs1

and 2a). Routinely collected programme data will also contribute to understanding teacher inputs in the Theory of Change, including attendance at CPD, use of resources and additional consultancy, and teacher outputs relating to planning. This data will be collected for all intervention schools and shared with the evaluation team at the end of the intervention period (Summer 2026).

Teacher surveys. An online teacher survey for all Year 2 and Year 5 teachers in intervention and control schools will be administered using Qualtrics software (n=260-390).[6] At baseline (prior to randomisation) the survey will be designed to understand teacher, school and class contexts and usual practice (IPERQ5). At endline, the survey will be designed to assess any changes in the control condition during the trial year (IPERQ5; monitoring of the control condition) and for intervention schools to explore the following: experience of training (IPERQ1), implementation and variability (IPERQ2), teacher knowledge and confidence (IPERQ3) and potential impact on pupils (IPERQ4). The survey will also enable the researchers to understand whether or not the teacher outputs, short-term teacher outcomes, pupil's inputs and outputs and teachers' perceptions of pupil outcomes, as detailed in the Theory of Change, are in place.

A measure of teacher self-efficacy in teaching writing will also be embedded in this survey at baseline and endline, as evidence suggests that teacher's writing self-efficacy impacts on the quality of teaching in writing (Abbott et al., 2023) and higher pupil writing outcomes (Tschannen-Moran & Johnson, 2011). This will be the **Preservice Teachers' Self-efficacy for Writing Instruction subscale of The Preservice Teacher Self-Efficacy for Writing Inventory (PTSWI)** (Hodges, Landau Wright & McTigue, 2021). The subscale has 11 items (e.g. 'I feel adequately prepared to teach writing') which have wider applicability beyond preservice teachers. Tests of validity and reliability were found to be aligned with wider research on writing (Cronbach's α=0.83). This was, however, developed with a US population, and so will be anglicised where necessary. Results will be used to answer IPERQ3 relating to teacher confidence and quality of implementation and assess the extent to which the associated teacher outcomes in the Theory of Change are in place. Where schools are unsure of the members of staff who will be teaching Year 2 and Year 5 classes in the academic year 2025/2026 the current Year 2 and Year 5 teachers will be requested to complete the survey and additional checking will be conducted prior to the end of the summer term (2025) and at the beginning of the autumn term (2025) to ensure an up-to-date list of Year 2 and Year 5 teachers participating in the trial. Any additional teachers will also be contacted at this point and requested to complete the baseline teacher survey. Emails and reminder phone calls will be used to encourage survey completion.

Sub-sample qualitative data collection. A sub-sample of intervention schools (n=10) will be selected to collect richer, more qualitative data particularly relating to implementation of Writing Roots in Years 2 and 5. They will allow for a more detailed understanding of whether or not the Theory of Change is working as expected and any potential mediators and moderators influencing variability of implementation. Within each of these sub-sample schools the following data will be collected:

- A Writing Roots **lesson observation** will be conducted in either a Year 2 or a Year 5 classroom per school (in order to minimise burden on schools; five Year 2 and five Year 5 lesson observations in total). The lesson plans and resources provided by

[6] Assuming an average of 3 teachers per school.

Literacy Tree will be used to form a structured framework for the lesson observations and they will be designed around the key elements of the programme identified in the programme Theory of Change, in particular, the pedagogical approach and key components of the programme which indicate fidelity (including teacher modelling, 'discovery points', discussion and debate) as well as the wider classroom environment (e.g. a working wall, published pupil work displayed). The observations will assist in answering IPERQ2b and c, relating to implementation fidelity and variability.

- A **semi-structured teacher interview** will also take place with the observed lesson teacher (five Year 2 teachers; five Year 5 teachers), ideally after the classroom observation. The interview schedule will be designed to cover the observed lesson, adaptations and potential (actual or perceived) barriers to implementation (IPERQ2b, c), pupil engagement and the perceived value of Writing Roots for different sub-sets of pupils including pupils receiving FSM and SEND (IPERQ4), and views on training and support (IPERQ1 and IPERQ3).
- A **pupil focus group** with either Year 2 (n=5) or Year 5 (n=5) pupils (to minimise burden on individual schools). Each focus group would have no more than four to five pupils and last a maximum of 20 minutes. Using their current Writing Roots text, the focus groups will explore pupils' engagement with texts, reading and writing for pleasure, understanding of writing conventions, audience and purpose (IPERQ4).
- Pupils involved in the pupil focus groups will also be asked to bring copies of their current writing books. The researcher/s will **photograph 4-5 pages of writing from exercise books** per pupil from the beginning of the academic year and the time of the focus group (Spring 2026). Where possible a piece of extended writing similar to that undertaken for the IE will be selected at both timepoints. This data will be used to understand more thoroughly the programmes layering approach, and the extent to which pupils are extending their writing as a result of Writing Roots (IPERQ4).

This qualitative sub-sample of schools will be selected based on school size, pupil demographics and geographical location. If there are a number of schools in the final main trial sample with mixed year group teaching this will also be taken into account when sampling for the qualitative sub-sample. One visit to a sub-sample school will be conducted by two researchers to ensure inter-observer reliability and a shared understanding of the observation schedule, as well as consistency in conducting interviews and focus groups. Burden to participating schools will be minimised by conducting as much of the data collection as possible in one visit, and by only focusing on one year group per school.

Interviews with a member of the Senior Leadership Team (SLT). To supplement the qualitative data collected in the sub-sample of schools we will conduct semi-structured interviews with a member of the SLT in ten intervention schools. These will be designed to provide rich contextual data relating to whole-school implementation and perceived impact of Literacy Tree on school culture (IPERQ2a), potential variability by year group and/or pupil factors (IPERQ2c). Given the importance to schools of national assessments (in Years 1, 2 and 6) these will also be discussed in the context of implementation and potential variability in the Writing Roots approach. In addition, the interviews will cover school engagement with training, in particular the extent to which schools are able to release teachers for training (IPERQ1a), barriers and facilitators to attendance at training (IPERQ1b), and levels of support (IPERQ2c) as well as potential unexpected consequences experienced as a result of the programme (IPERQ6). The interviews with members of SLT will contribute to a whole-school understanding of the programme Theory of Change. It is intended that the sample will consist of five members of SLT from the qualitative sub-sample schools to allow for triangulation and

five additional intervention schools to widen the overall sample for the qualitative data. SLTs will be selected from schools based on school size, pupil demographics and geographical location. It is anticipated that these interviews will occur online to avoid overburdening members of SLT during the school day.

Control school visits. In order to understand the control condition more fully, the researchers will visit a small sample of control schools (n=4). Each visit will consist of a **writing lesson observation**, to identify key elements of reading and writing pedagogy and curriculum as opposed to key elements of Writing Roots, to understand the extent to which the teaching of writing and wider literacy practices in control schools differs from the intervention (IPERQ5). The observations will be conducted prior to the intervention school visits so that normal practice can be established. Two observations will be of a Year 2 writing lesson and two will be of a Year 5 writing lesson. The observation schedule will consist of a checklist designed around the researchers' prior experience of observing writing lessons in primary schools (e.g. Grammar for Writing; Tracey et al., 2019) and include key aspects of writing instruction such as teaching of vocabulary and writing conventions, use of texts, and spelling. These will be supplemented by researcher notes. Alongside this we anticipate also conducting **semi-structured interviews** with four control group teachers (ideally those teachers who have had their lesson observed) to understand the observed lessons and usual practice more fully (two Year 2 teachers; two Year 5 teachers). Schools will be purposively selected based on school size, pupil demographics and geographical location. One visit to a control school will be conducted by two researchers to ensure inter-observer reliability and a shared understanding of the observation schedule as well as consistency in conducting interviews. Burden to participating schools will be minimised by conducting as much of the data collection as possible in one visit, and by only focusing on one year group per school (i.e. visits will focus on Year 2 in two control schools and on Year 5 in the remaining two control schools).

In the follow-up study an **online survey** will be administered to all Year 2 and Year 5 teachers in intervention schools via Qualtrics in Summer 2027. They have been chosen as they were the main focus of the evaluation during the main trial and we will have more existing information relating to their implementation during the intervention period (i.e. from surveys, interviews and observations). This survey will be designed to explore the extent to which the programme is embedded within schools in the year after the intervention training. It will cover: implementation and variability, and whether or not schools have implemented any additional literacy programmes since the end of the main trial (IPERQ7); teacher knowledge and confidence, and any additional support needed for continued implementation (including due to teacher turnover and potential cascading of the programme training) (IPERQ8); and potential impact on pupils (IPERQ9). As with the baseline and endline surveys in Year 1, this follow-up survey will also have the **Preservice Teachers' Self-efficacy for Writing Instruction subscale of the PTSWI** embedded to enable the researchers to explore potential changes in teacher self-efficacy over time (IPERQ8). We anticipate the sample will involve 58 schools and 116 teachers.[7] Schools will be contacted at the beginning of the academic year to capture any changes of teacher in Year 2 and Year 5 for the academic year 2026-2027 and to ensure contact details are up to date. Emails and reminder phone calls will be used to encourage survey completion during the Summer Term 2027.

[7] Assuming low levels of attrition from the main trial and from the extension study (5% at each timepoint) and a conservative one teacher per year group per intervention school.

Analysis

Throughout the trial year, quantitative data from the teacher surveys will be downloaded from Qualtrics into SPSS and cleaned. Baseline data will be analysed descriptively to provide information relating to the teacher and school context and usual literacy teaching (IPERQ5). Endline data from intervention schools will be analysed descriptively to provide understanding of experiences of training (IPERQ1), implementation (IPERQ2) and perceived outcomes from the programme (IPERQ4). Additional analysis of baseline and endline responses will allow for comparisons over time and between conditions, in particular relating to changes in teacher knowledge and confidence, and self-efficacy (as measured by the Self-efficacy for Writing Instruction subscale of the PTSWI) (IPERQ3), changes in usual practice and any potential unexpected consequences, such as changes in the control condition during the trial year (IPERQ5 & IPERQ6). Additional analysis will be conducted to compare responses between year groups (i.e. Year 2 and Year 5 teachers) to assess variability (IPERQ2). Where teachers complete the survey at different timepoints (i.e. during the summer or autumn term 2025) an exploratory analysis will be conducted to assess for differences in responses.

Observation checklists for the INSET events and 'Teach Through a Text' training sessions (IPERQ1) and for the lesson observations (in control and intervention schools) (IPERQ2 & IPERQ5) will be inputted into SPSS and cleaned. They will be analysed descriptively using frequency counts. Attendance at training and the routinely collected programme data will be uploaded into SPSS and analysed descriptively at both the individual and the school level (IPERQ1 & IPERQ2). The photographs of the writing samples from pupils' textbooks will be marked using the WAM marking structure (Dunsmuir et al., 2015). This will include all domains included in the WAM, including those outside of the IE, in order to explore potential unexpected consequences (IPERQ6): handwriting, spelling, punctuation, sentence structure and grammar, vocabulary, organisation and overall structure, and ideas. The findings will be reported descriptively for each sub-scale as well as the overall total score, with comparisons made between the two timepoints (IPERQ4). Quality assurance checks will be put in place to ensure that all statistical analysis (of surveys, observation checklists and writing sub-sample analysis) can be replicated.

Interviews (with teachers, trainers and members of schools' SLT) and pupil focus groups will be transcribed by a University of Leeds approved transcription service with appropriate data sharing agreements in place. Researcher notes from launch event and planning surgery observations and from intervention and control lesson observations will be typed into Word documents. Open text responses from the surveys will be downloaded from Qualtrics into Excel. All qualitative data will be uploaded and coded in NVivo using a deductive approach (Vanover, Saldaña & Mihas, 2022) alongside a thematic analysis (Braun & Clarke, 2021). This will allow the researchers to answer the research questions and to systematically interrogate the programme mechanisms as outlined in the Theory of Change whilst allowing any underlying themes to emerge from the data itself. Qualitative analysis will thus contribute to answering all the IPE research questions, including IPERQ6 relating to unexpected consequences. Two researchers will conduct the qualitative analysis with inter-rater reliability checks established.

Where more than one data source is used to answer a research question the findings will be triangulated and synthesised with supporting evidence from the appropriate data source to ensure trustworthiness of the findings. In addition, findings from the IPE and IE will be triangulated at a synthesis workshop so that the IPE and IE analysis can be used together to

fully understand outcomes and to assess the extent to which the programme Theory of Change is working as expected. Table 5 presents an overview of the methods, sampling and analysis for the main trial year of the evaluation.

Table 5: IPE methods overview – main trial

IPE dimension	RQ addressed	Research methods	Data collection methods	Sample size and sampling criteria	Data analysis methods
Fidelity	IPERQ1, IPERQ2	Observation	Training observations	2 INSET launch events (n=2 intervention schools) 6 'Teach Through a Text' online training sessions (Year 2 n=2; Year 5 n=2; Year 1 n=1; Year 6 n=1)	Count analysis Deductive coding, thematic analysis
		Interview	Online interviews with trainers	3 trainers	Deductive coding, thematic analysis
		Routine data	Training and support data, engagement with on-line resources	65 intervention schools	Descriptive statistics
		Survey (endline, intervention)	Online Teacher surveys	195 intervention teachers (Year 2 and Year 6) (assumption av. 3 teachers per school)	Descriptive statistics, deductive coding, thematic analysis
		Observation	Intervention lesson observations	10 lesson observations in intervention schools (Year 2 n=2; Year 5 n=2)	Count analysis Deductive coding, thematic analysis

		Interview	Face to face interviews with teachers	10 interviews with intervention teachers (Year 2 n=5; Year 5 n=5)	Deductive coding, thematic analysis
		Interview	Online interviews with SLTs	10 interviews with SLT members in intervention schools	Deductive coding, thematic analysis
Dosage	IPERQ1	Routine data	Training attendance data	65 intervention schools	
Responsiveness	IPERQ1, IPERQ3, IPERQ4	Observation	Training observations	2 INSET launch events (n=2 intervention schools) 6 'Teach Through a Text' online training sessions (Year 2 n=2; Year 5 n=2; Year 1 n=1; Year 6 n=1)	Count analysis Deductive coding, thematic analysis
		Interview	Online interviews with trainers	3 trainers	Deductive coding, thematic analysis
		Survey (endline, intervention)	Online Teacher surveys	195 intervention teachers (Year 2 and Year 6) (assumption av. 3 teachers per school)	Descriptive statistics, deductive coding, thematic analysis
		Interview	Face to face interviews with teachers	10 interviews with intervention teachers (Year	Deductive coding, thematic analysis

				2 n=5; Year 5 n=5)	
		Interview	Online interviews with SLTs	10 interviews with SLT members in intervention schools	Deductive coding, thematic analysis
		Focus group	Face to face pupil focus groups	10 focus groups, 4-5 pupils per focus group, intervention schools (Year 2 n=5; Year 5 n=5)	Deductive coding, thematic analysis
		Writing samples	Pupil exercise books writing samples	80-100 writing samples, 40-50 pupils (4-5 pages per pupil taken from two timepoints: Year 2 n=20-25 pupils; Year 5 n=20-25 pupils)	Descriptive analysis, t-tests
Quality	IPERQ2, IPERQ3	Observation	Lesson observations	10 lesson observations in intervention schools (Year 2 n=2; Year 5 n=2)	Count analysis Deductive coding, thematic analysis
		Interview	Face to face interviews with teachers	10 interviews with intervention teachers (Year 2 n=5; Year 5 n=5)	Deductive coding, thematic analysis
		Interview	Online interviews with SLTs	10 interviews with members of SLT from 10 intervention schools (one	Deductive coding, thematic analysis

				from each school)	
		Survey (endline, intervention)	Online Teacher surveys	195 intervention teachers (Year 2 and Year 6) (assumption av. 3 teachers per school)	Descriptive statistics, deductive coding, thematic analysis
Reach	IPERQ1 IPERQ2, IPERQ4	Observation	Training observations	2 INSET launch events (n=2 intervention schools) 6 'Teach Through a Text' online training sessions (Year 2 n=2; Year 5 n=2; Year 1 n=1; Year 6 n=1)	Count analysis Deductive coding, thematic analysis
		Interview	Online interviews with trainers	3 trainers	Deductive coding, thematic analysis
		Survey (endline, intervention)	Online Teacher surveys	195 intervention teachers (Year 2 and Year 6) (assumption av. 3 teachers per school)	Descriptive statistics, deductive coding, thematic analysis
		Interview	Face to face interviews with teachers	10 interviews with intervention teachers (Year 2 n=5; Year 5 n=5)	Deductive coding, thematic analysis
		Interview	Online interviews with SLTs	10 interviews with SLT members in	Deductive coding, thematic analysis

				intervention schools	
		Focus group	Face to face pupil focus groups	10 focus groups, 4-5 pupils per focus group, intervention schools (Year 2 n=5; Year 5 n=5)	Deductive coding, thematic analysis
Adaptations	IPERQ2	Observation	Lesson observations	10 lesson observations in intervention schools (Year 2 n=2; Year 5 n=2)	Count analysis Deductive coding, thematic analysis
		Interview	Face to face interviews with teachers	10 interviews with intervention teachers (Year 2 n=5; Year 5 n=5)	Deductive coding, thematic analysis
		Interview	Online interviews with SLTs	10 interviews with SLT members in intervention schools	Deductive coding, thematic analysis
		Survey (endline, intervention)	Online Teacher surveys	195 intervention teachers (Year 2 and Year 6) (assumption av. 3 teachers per school)	Descriptive statistics, deductive coding, thematic analysis
		Interview	Online interviews with trainers	3 trainers	Deductive coding, thematic analysis

Monitoring of the control condition	IPERQ5, IPERQ6	Survey (baseline - intervention and control/ endline - control)	Online Teacher surveys	Baseline:390 teachers, intervention and control (Year 2 and Year 6) (assumption av. 3 teachers per school) Endline: 195 teachers, control (Year 2 and Year 6) (assumption av. 3 teachers per school)	Descriptive statistics, deductive coding, thematic analysis
		Observation	Control school lesson observations	4 lesson observations, 1 from each of 4 control schools (Year 2 n=2; Year 5 n=2)	Count analysis Deductive coding, thematic analysis
		Interview	Face to face interviews with teachers	4 control teachers, 1 from each of 4 control schools (Year 2 n=2; Year 5 n=2)	Deductive coding, thematic analysis

In the follow-up study, quantitative data from the follow-up intervention teacher surveys will be downloaded from Qualtrics into SPSS and cleaned. They will be analysed descriptively and, where possible, they will be matched with endline teacher surveys to assess change over time. Additional analysis may also be conducted to compare year groups (Year 2 and Year 5 teacher responses) to assess variability (IPERQ7). A further synthesis workshop will be held to triangulate the findings with the main trial year findings and the longitudinal analysis from the NPD. Table 6 presents an overview of the methods, sampling and analysis for the follow-up study.

Table 6: IPE methods overview – follow-up study

IPE dimension	RQ addressed	Research methods	Data collection methods	Sample size and sampling criteria	Data analysis methods
Fidelity	IPERQ7	Survey (follow-up)	On-line Intervention teacher survey	116 teachers (Year 2 and Year 5 teachers, intervention only)	Descriptive statistics, deductive coding, thematic analysis
Responsiveness	IPERQ8, IPERQ9	Survey (follow-up)	Intervention teacher survey	116 teachers (Year 2 and Year 5 teachers, intervention only)	Descriptive statistics, deductive coding, thematic analysis
Quality	IPERQ7, IPERQ8	Survey (follow-up)	Intervention teacher survey	116 teachers (Year 2 and Year 5 teachers, intervention only)	Descriptive statistics, deductive coding, thematic analysis
Reach	IPERQ7 IPRQ9	Survey (follow-up)	Intervention teacher survey	116 teachers (Year 2 and Year 5 teachers, intervention only)	Descriptive statistics, deductive coding, thematic analysis

Cost evaluation design

Costs will be evaluated using data from endline teacher surveys, with business as usual costs for literacy delivery serving as the counterfactual for the cost of Writing Roots, which may include the direct and indirect costs of running similar programmes. We note that 'business as usual' is likely to vary substantially across settings.

As per EEF cost evaluation guidance (2023), we will calculate the total cost to the setting and the cost per-pupil-per-school-year across three years of delivery. This evaluation will take into account the direct cost of Writing Roots (i.e. the fee) and the indirect costs, which may include, but will not be limited to, costs related to teacher training and teacher cover; costs related to materials and equipment and materials participating schools may need to produce; and unexpected and hidden costs, which are not anticipated by the delivery team.

Ethics and registration

The trial will be registered on the International Standard Randomised Controlled Trial Number (ISRCTN) registry, which is used to describe randomised controlled trials (RCTs) and efficacy

trials at inception. Once registered, this protocol will be updated with the assigned registration number.

The ethics and registration processes are in accordance with the ethics policies adopted by RAND Europe and the University of Leeds. The trial received ethical approval from RAND Europe's internal review board on 24/02/2025 (ref: 022807.021). It has also received ethical approval from the University of Leeds Cross-Faculty Research Ethics Committee (Faculties of Business, Environment & Social Sciences) (Ref: 2583; 31/03/2025).

Prior to pupils' data being sent to the delivery team, parents will be sent information sheets and withdrawal forms by the setting and will have the opportunity to return these. Parents can withdraw their children at any time from the data collection activities. Parents will be given two weeks between when information sheets are sent out and when pupil data is collected from settings, to allow parents to withdraw children from the evaluation before any data is collected. If parents choose to withdraw their children from data collection later on, their data will not be collected or will be deleted, as appropriate.

RAND Europe will collect consent forms for all practitioners, managers and trainers that participate in an interview. The front page for each online survey will contain a privacy notice informing respondents that participation in the survey is entirely voluntary. The consent form in the survey will be built into the data collection tool so that those moving past a certain page (following the privacy notice and information on the research) will have given consent for the data to be used in the research.

None of the evaluation team has any conflicts of interest and all members of the study team have approved this protocol prior to publication.

Data protection

RAND and University of Leeds are committed to ensuring that all research we undertake is ethical and complies with the highest standards to ensure safety for participants and researchers. RAND Europe and the University of Leeds will act as data controllers and will be the main point of contact for any matters relating to the protection of all personal data once the recruitment is completed and will make decisions about how and what personal data is used in the evaluation. We will establish how data will be shared and used, generating a data flow diagram to support all parties to understand roles and responsibilities, which will feed into a data privacy impact assessment (DPIA) in accordance with ICO guidance on processing personal, identifiable data. We, alongside our dedicated Data Protection Officers, have extensive experience developing appropriate privacy notices, withdrawal forms, and information sheets for parents and schools, as well as data sharing agreements to ensure data is managed safely.

Data will be handled in accordance with the General Data Protection Regulations (GDPR). Personal data will be processed under Article 6 Section (e) of the GDPR ('Tasks carried out in the public interest') as the research is being conducted to support education provision in the UK (and, if applicable, Special Category data under Article 9(2)(j)). Our researchers are accredited by the Office for National Statistics to use data from the NPD.

Personnel

Delivery team: Literacy Tree

Project manager: Anthony Legon

Head of consultancy: Lynn Sear

Recruitment and communications manager: Bronte Larsen-Disney

Recruitment and accounts manager: Jules Moody

School and research lead: Hannah Baker

Evaluation team: RAND Europe and University of Leeds

Principal investigator and project leader: Elena Rosa Speciani

Project Manager: Rebecca Mead

AI marking: Andy Skelton, James Merewood, Elaine Wang

Impact evaluation lead: Vincent Melnikov

Implementation and process evaluation and data collection: Louise Tracey, Erin Dysart, Paula Clarke

Risks

Risks	Assessment	Mitigation strategy	Impact post-mitigation
Under-recruitment	Likelihood: Moderate Impact: High	Regular dialogue over recruitment issues and ensuring that the design poses minimal burden to schools. Timelines to be discussed and agreed well in advance to ensure adequate time for all activities.	Low
Attrition	Likelihood: Moderate Impact: High	Schools to be given clear information about what participation entails before signing up. Keeping test burden low to maximise participation and reduce dropout. Mop-up testing to reduce pupil level attrition. Factoring potential for attrition into power calculations for determining sample size.	Moderate
Not finding an appropriate	Likelihood: Moderate	Use set-up period to understand key outcomes in ToC, drawing upon the team's extensive experience of	Low

writing assessment	Impact: High	assessing writing to propose and select appropriate measures. Using the pre-randomisation period to propose, pilot, and check the reliability and validity of a new writing prompt for WAM to fit the needs of the evaluation.	
AI tool not reliable	Likelihood: Moderate Impact: High	Initial testing conducted prior to proposal shows promise. Pilot process and establish interrater reliability with human markers. Multiple opportunities to test and adapt during pilot phase. Pilot well in advance of the trial so human markers can be recruited and trained if needed.	Low
Low participation rates in data collection (testing and IPE)	Likelihood: Moderate Impact: High	Providing ample time for data collection, with real-time monitoring of response rates to allow for sending of targeted reminders. Piloting measures and data collection tools to understand how to reduce burden on schools.	Low
Small number of FSM pupils for analysis	Likelihood: Moderate Impact: Moderate	Recruiting schools in high-deprivation areas to support FSM sub-group analysis in both the IE and IPE. Focusing mop-up testing in schools with high numbers of FSM pupils.	Low
Evaluation team members absence or turnover	Likelihood: Moderate Impact: Low	Supplementing the team with experienced evaluation researchers from the larger RE/UL pool. Three-month notice period for all staff, to allow sufficient handover time.	Low

Timeline

Activity	Dates	Staff responsible/ leading
Expression of interest deadline	July 2025	Literacy Tree
MOU to be signed	July 2025	RAND Europe
Baseline assessment writing task & baseline Year 2 and Year 5 teacher on-line surveys	June-July 2025	University of Leeds

Randomisation of settings into intervention and control groups. Inform settings of randomisation allocation	July 2025	RAND Europe
Delivery of INSET online for whole school Year 1-6 (intervention schools only)	September 2025	Literacy Tree
Writing Roots programme delivery. Teach Through a Text year group specific training October and January (intervention schools only)	September 2025 – June 2026	Literacy Tree
Selected school visits by University of Leeds	February – April 2026	University of Leeds
Endline assessment & endline Year 2 and Year 5 teacher online surveys	June – July 2026	University of Leeds
Intervention schools offered continued Literacy Tree membership	July 2026	Literacy Tree
Control group offered Literacy Tree membership for next academic year	July 2026	Literacy Tree
Initial report on impact of Writing Roots on Year 2 and Year 5 writing	Spring 2027	RAND Europe & University of Leeds
Follow-up online intervention teacher survey	June 2027	University of Leeds
Final findings published (Y6 SATs and teacher survey)	Spring 2028	University of Leeds

References

Abbott, J., Hodges, T., Dismuke, S., Landau Wright, K. & Schweiker, C. (2022). Interventions to Improve Teacher Self-Efficacy Beliefs about Writing and Writing Instruction: Lessons Learned and Areas for Exploration. *Teaching/Writing: The Journal of Writing Teacher Education*: Vol. 11: Iss. 3, Article 7. Available at: https://scholarworks.wmich.edu/wte/vol11/iss3/7

Anders, J., Shure, N., Wyse, D., Barnard, M., Abdi, F. & Frerichs, J. (2021). *Power of Pictures: Evaluation Report*. London: Education Endowment Foundation. Available from: https://d2tic4wvo1iusb.cloudfront.net/production/documents/pages/projects/Power_of_Pictures_Evaluation_Report_Final.pdf?v=1744107788

Anders, J. Shure, N., Wyse, D., Bohling, K., Sutherland, A., Barnard, M. & Frerichs, J. (2021). *Learning About Culture. Overarching Evaluators' Report.* London: Education Endowment Foundation. Available from: https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/methodological-research-and-innovations/learning_about_culture_overarching_evaluators_report.pdf?v=1723691903

Bandura, A. (1997). Self-efficacy: The exercise of control. New York: W.H. Freeman.

Berninger, V. W., Abbott, R. D., Abbott, S. P., Graham, S., & Richards, T. (2002). Writing and Reading: Connections Between Language by Hand and Language by Eye. *Journal of Learning Disabilities*, *35*(1), 39–56. https://doi.org/10.1177/002221940203500104

Berninger, V. W., & Winn, W. D. (2006). Implications of advancements in understanding executive function and writing for education. Journal of Learning Disabilities, 39(4), 320–328.

Braun, V. & Clarke, V. (2021). *Thematic Analysis: A Practical Guide.* Sage.

Brown, E.R., Subosa, M., Merewood, J., Senator, B., Hutton, E. & Reed, M. (2022). *Independent evaluation of the Lexia® Core5® Reading Programme: A cluster randomised controlled trial. Evaluation Protocol.* London: Education Endowment Foundation. Available from: https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/Evaluation-protocol-Lexia.pdf?v=1727267593

Bruning, R., Dempsey, M., Kauffman, D. F., McKim, C., & Zumbrunn, S. (2013). Examining Dimensions of Self-Efficacy for Writing. *Journal of Educational Psychology*, *105*(1), 25–38. https://doi.org/10.1037/a0029692

Bulmer, M. G. (1979). *Principles of Statistics*. Dover Publications.

Campbell, T., Cooper, K., & Fowler, J. (2025). *Who has been registered for Free School Meals and pupil premium in the National Pupil Database?* Education Policy Institute. Available: https://epi.org.uk/wp-content/uploads/2025/03/FSM-report-March-2025_PDF.pdf

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Lawrence Erlbaum Associates.

Creswell, J. W., & Creswell, J. D. (2017). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Sage Publications.

Darling-Hammond, L., Hyler, M. E., Gardner, M. (2017). Effective Teacher Professional Development. Palo Alto, CA: Learning Policy Institute. Available from: https://learningpolicyinstitute.org/product/effective-teacher-professional-development-report

DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods*, 2(3), 292-307.

Department for Education (DfE). (2019). Improving cultural capital in schools. London: DfE.

Department for Education (DfE). (2021). *Education recovery strategy*. London: DfE.

Dunsmuir, S., Kyriacou, M., Batuwitage, S., Hinson, E., Ingram, V., & O'Sullivan, S. (2015). An evaluation of the Writing Assessment Measure (WAM) for children's narrative writing. *Assessing Writing*, *23*, 1–18. https://doi.org/10.1016/j.asw.2014.08.001

EEF (2020). *Improving Literacy in Key Stage 1*. Second Edition. London: Education Endowment Foundation. Available from: https://d2tic4wvo1iusb.cloudfront.net/production/eef-guidance-reports/literacy-ks-1/Literacy_KS1_Guidance_Report_2020.pdf?v=1722359203

EEF (2021). *Improving Literacy in Key Stage 2*. Second Edition. London: Education Endowment Foundation. Available from: https://d2tic4wvo1iusb.cloudfront.net/production/eef-guidance-reports/literacy-ks2/EEF-Improving-literacy-in-key-stage-2-report-Second-edition.pdf?v=1744040067

EEF (2022). *Statistical analysis guidance for EEF evaluations.* London: Education Endowment Foundation. Available from: https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/evaluation-design/EEF-Analysis-Guidance-Website-Version-2022.14.11.pdf?v=1744040298

EEF (2023) *Cost evaluation guidance for EEF evaluations*. London: Education Endowment Foundation. Available from: https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/evaluation-design/Cost-Evaluation-Guidance-Feb_2023.pdf?v=1727947494

Finlayson, K., and McCrudden, M. T. (2019). "Teacher-Implemented Writing Instruction for Elementary Students: A Literature Review," *Reading & Writing Quarterly*, Vol. 36, No. 1. https://doi.org/10.1080/10573569.2019.1604278

Golparvar, A. E., & Khafi, A. (2021). The role of L12 writing self-efficacy in integrated writing strategy use and performance. Assessing Writing, 47. https://doi.org/10.1016/j.asw.2020.100504.

Graham, S., McKeown, D., Kiuhara, S., & Harris, K. R. (2012). A meta-analysis of writing instruction for students in the elementary grades. *Journal of Educational Psychology, 104*(4), 879–896. https://doi.org/10.1037/a0029185

Graham, S., and Perin, D. (2007a). "A Meta-Analysis of Writing Instruction for Adolescent Students," *Journal of Educational Psychology*, Vol. 99, No. 3. https://doi.org/10.1037/0022-0663.99.3.445

Graham, S., & Perin, D. (2007b). *Writing next: Effective strategies to improve writing of adolescents in middle and high schools – A report to Carnegie Corporation of New York.*

Washington, DC: Alliance for Excellent Education. Available from: https://media.carnegie.org/filer_public/3c/f5/3cf58727-34f4-4140-a014-723a00ac56f7/ccny_report_2007_writing.pdf

Goswami, U. (2015). *Children's Cognitive Development and Learning*. York : Cambridge Primary Review Trust. Available from: https://bristolearlyyearsresearch.org.uk/wp-content/uploads/2017/06/3._COMPLETE-REPORT-Goswami-Childrens-Cognitive-Development-and-Learning.pdf

Gutman, L., M. & Schoon, I. (2013). *The impact of non-cognitive skills on outcomes for young people.* London: Education Endowment Foundation. Available from: https://discovery.ucl.ac.uk/id/eprint/10125763/1/Gutman_Schoon_%202013%20Non-cognitive_skills_literature_review_.pdf

Harris, A., & Jones, M. (2017). Leading educational change and improvement at scale: some inconvenient truths about system performance. *International Journal of Leadership in Education*, *20*(5), 632–641. https://doi.org/10.1080/13603124.2016.1274786

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Ransdell (Eds.), The science of writing: Theories, methods, individual differences, and applications (pp. 1–27). Mahwah, NJ: Erlbaum.

Hodges, T. S., Wright, K. L., & McTigue, E. M. (2021). The Preservice Teacher Self-Efficacy for Writing Inventory (PTSWI): A tool for measuring beliefs about writing. *Assessing Writing*, *49*, 100545-. https://doi.org/10.1016/j.asw.2021.100545

Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R. & Kerr, K. (2019). Implementation and process evaluation (IPE) for interventions in education settings: An introductory handbook. Education Endowment Foundation.

Kennedy, M. M. (2016). How Does Professional Development Improve Teaching? *Review of Educational Research*, *86*(4), 945–980. https://doi.org/10.3102/0034654315626800

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174.

Li, Z., Hassan, N. C., & Jalil, H. A.. (2023). The Effectiveness of Face-to-Face versus Online Delivery of Continuing Professional Development for Science Teachers: A Systematic Review. *Education Sciences*, *13*(12), 1251. https://doi.org/10.3390/educsci13121251

Lord, P., Styles, B., Morrison, J., White, R., Andrade, J., Bamford, S., Lushey, C., Lucas, M. & Smith, R. (2018). *Families and Schools Together (FAST)*. London: Education Endowment Foundation. Available from: https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/FAST.pdf?v=1727258054

Loxley, A., Johnston, K., Murchan, D., Fitzgerald, H., & Quinn, M. (2007). The role of whole-school contexts in shaping the experiences and outcomes associated with professional development. *Journal of In-service Education*, *33*(3), 265–285. https://doi.org/10.1080/13674580701487034

Loxley, A., Johnston, K., Murchan, D., Fitzgerald, H., & Quinn, M. (2007). Whole-school action research for development in schools. Educational Action Research, 15(3), 337–352.

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). Intelligence Unleashed: An Argument for AI in Education. Pearson.

Myhill, D., Jones, S., & Watson, A. (2012). Grammar matters: How teachers' grammatical knowledge impacts on the teaching of writing. Teaching and Teacher Education, 28(5), 759–767.

National Literacy Trust (2017). *Primary Annual Literacy School Review*. London: National Literacy Trust. Available from: https://cdn.literacytrust.org.uk/media/documents/2017_08_25_memb_resource_-_Primary_ALSR_4_Writing_culture_poster.pdf

Ofsted. (2022). Annual report of Her Majesty's Chief Inspector of Education, Children's Services and Skills 2021/22. London: Ofsted.

Ofsted (2022). *Research review series: English*. GOV.UK. Available from: https://www.gov.uk/government/publications/curriculum-research-review-series-english/curriculum-research-review-series-english (Accessed 8 April 2025).

O'Leary, C., Akhtar, K. & Marley, B. (2024). *The benefits of including diverse children's books in a classroom*. Book Trust. Available from: https://www.booktrust.org.uk/news-and-features/features/2024/march/the-benefits-of-including-diverse-childrens-books-in-a-classroom/#:~:text=%22Representation%20is%20one%20of%20a,in%20the%20stories%20they%20read (Accessed 8 April 2025).

O'Leary, J., Smith, R., & Jones, T. (2024). Diversity in classroom texts and pupil engagement: A longitudinal study. British Journal of Educational Research, 40(1), 45–62.

OUP (2023). *Oxford Language Report 2023*. Oxford University Press: Education and Children's Books. Available from: https://global.oup.com/education/press/oxford-language-report-2023-4-in-10-pupils-have-fallen-behind/?region=international (Accessed 8 April 2025).

Ray, D., Muñoz, A., Zhang, M., Li, X., Chatterjee, N., Jacobson, L. P., & Lau, B. (2022). Meta-analysis under imbalance in measurement of confounders in cohort studies using only summary-level data. *BMC Medical Research Methodology*, *22*(1). https://doi.org/10.1186/s12874-022-01614-9

Rogers, L. A., & Graham, S. (2008). A meta-analysis of single subject design writing intervention research.*Journal of Educational Psychology, 100*(4), 879–906. https://doi.org/10.1037/0022-0663.100.4.879

Saldana, J., Vanover, C. & Mihas, P. (2021). *Analyzing and Interpreting Qualitative Research: After the Interview*. Digital Commons @ University of South Florida.

Slavin, R. E., Lake, C., Davis, S., & Madden, N. (2019). Effective programs for improving reading and writing achievement. *Educational Research Review*, 27, 1–20.

Shermis, M.D., & Burstein, J. (Eds.). (2013). Handbook of Automated Essay Evaluation: Current Applications and New Directions (1st ed.). Routledge. https://doi.org/10.4324/9780203122761

Singh, A., Uwimpuhwe, G., Vallis, D., Akhter, N., Coolen-Maturi, T., Higgins, S., Einbeck, J., Culliney, M., & Demack. S. (2023). Improving Power Calculation in Education Trials.

Slavin, R. E., Lake, C., & Groff, C. (2009). Effective programs in elementary mathematics: A best-evidence synthesis. *Review of Educational Research, 79(3), 1391–1466*.

Slavin, R. E., Lake, C., Inns, A., Baye, E., Dachet, D. & Haslam, J. (2019). *A Quantitative Synthesis of Writing Approaches in Years 3 to 13*. London: Education Endowment Foundation. Available from: https://files.eric.ed.gov/fulltext/ED612219.pdf

Spandel, V. (2006). Creating Writers Through 6-Trait Writing Assessment and Instruction. Pearson.

Stone, G., Andrade, J., Martin, K. & Styles, B. (2022). *Helping Handwriting Shine: Evaluation Report*. London: Education Endowment Foundation. Available from: https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/Helping-Handwriting-Shine-Addendum-Report-Final.pdf?v=1727266174

Tschannen-Moran, M., & Johnson, D. (2011). Exploring literacy teachers' self-efficacy beliefs: Potential sources at play. Teaching and Teacher Education, 27(4), 751–761.

Tracey, L., Boehnke, J. R., Elliott, L., Thorley, K., Ellison, S., & Bowyer-Crane, C. A. (2019). Grammar for Writing: Evaluation report and executive summary. Education Endowment Foundation.

Van der Kleij, F., Feskens, R., & Eggen, T. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes: A meta-analysis. *Educational Research Review*, 14, 1–26.

Vanover, C., Saldaña, J. & Mihas, P. (2022). Analyzing and Interpreting Qualitative Research: After the Interview. (2022). India: SAGE Publications, Incorporated.

Vescio, V., Ross, D., & Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. *Teaching and Teacher Education*, *24*(1), 80–91. https://doi.org/10.1016/j.tate.2007.01.004

Williamson, D.M., Mislevy, R.J., & Bejar, I.I. (Eds.). (2006). Automated Scoring of Complex Tasks in Computer-Based Testing (1st ed.). Routledge. https://doi.org/10.4324/9780415963572