Education Endowment Foundation

Power of Pictures

Evaluation Report

September 2021

Authors:

Jake Anders, Nikki Shure, Dominic Wyse, Matthew Barnard, Faisa Abdi, Johanna Frerichs

UCL Institute of Education

THE BEHAVIOURAL INSIGHTS TEAM

The Education Endowment Foundation (EEF) is an independent grant-making charity dedicated to breaking the link between family income and educational achievement, ensuring that children from all backgrounds can fulfil their potential and make the most of their talents.

The EEF aims to raise the attainment of children facing disadvantage by:

- identifying promising educational innovations that address the needs of disadvantaged children in primary and secondary schools in England;
- evaluating these innovations to extend and secure the evidence on what works and can be made to work at scale; and
- encouraging schools, government, charities, and others to apply evidence and adopt innovations found to be effective.

The EEF was established in 2011 by the Sutton Trust as lead charity in partnership with Impetus (formerly Impetus Trust) and received a founding £125m grant from the Department for Education.

Together, the EEF and Sutton Trust are the government-designated What Works Centre for improving education outcomes for school-aged children.

For more information about the EEF or this report please contact:

Jonathan Kay

Education Endowment Foundation

5th Floor, Millbank Tower

21–24 Millbank

SW1P 4QP

0207 802 1653

jonathan.kay@eefoundation.org.uk

www.educationendowmentfoundation.org.uk

Contents

About the evaluator

The project was independently evaluated by a team from University College London (Dr Jake Anders, Dr Nikki Shure, Professor Dominic Wyse, Professor John Jerrim, Professor Gemma Moss, Professor Andrew Burn) and the Behavioural Insights Team (Dr Matt Barnard, Kim Bohling, Johanna Frerichs, Faisa Abdi).

The lead evaluator was Dr Jake Anders.

Contact details:

UCL Centre for Education Policy and Equalising Opportunities

Institute of Education

20 Bedford Way

London WC1H 0AL

Email: cepeo@ucl.ac.uk

Telephone: +44 (0)20 7612 6000

Executive summary

The project

Power of Pictures (PoP) is a programme that aims to raise children's reading and writing skills by enhancing teachers' understanding of the power of picturebooks and increasing teachers' comfort and ability in teaching using picturebooks. PoP is provided to English schools by the Centre for Literacy in Primary Education (PE).

PoP comprises of teacher training delivered directly from the author-illustrator of a picturebook that will be taught in schools. Pupils also meet PoP programme deliverers and the author-illustrator of the picturebook that they will study during a half day workshop, focusing on character design and the development of narratives. Teachers then support pupils to develop their own picturebook stories, inspired by the workshop. Schools taking part in the programme are required to commit at least four weeks of their English curriculum to PoP (equal to 20 lessons over two terms).

This efficacy two-armed randomised controlled trial saw 1264 pupils in 51 schools receive the programme, while 1410 pupils in 50 schools acted as a comparison group. The trial evaluated the impact of PoP on Year 5 pupils' writing skills, writing self-efficacy and writing creativity (ideation). Additional to the impact evaluation of this trial, an implementation and process evaluation (IPE) was used to understand how teachers experienced the PoP training, how teachers delivered PoP lessons, and to better understand the barriers and facilitators to implementation and delivering the programme with fidelity.

The trial started in November 2017 with programme delivery from October 2018 to June 2019. This evaluation was jointly funded by the EEF and the Royal Society of Arts.

Key conclusions

1. Pupils who received the Power of Pictures (PoP) programme had, on average, higher writing scores (equivalent to one month of additional progress) as compared to children in the control group. This is our best estimate of impact which has a moderate to high security rating. However, as with any study, there is uncertainty around the result: the possible impact of this programme on reading attainment ranges from one month less progress to positive effects of three additional months of progress.
2. Among children eligible for free school meals (FSM), those in schools that PoP was delivered in also made one additional month's progress. These results may have a lower security rating than the overall findings because of the smaller number of pupils in this group.
3. Children in PoP schools had higher writing self-efficacy and writing creativity (ideation) scores than those from schools in which the programme was not taught. These differences suggest that PoP may have a positive impact on these outcomes.
4. The visual element of this programme attracted learners who traditionally have difficulties engaging in literacy activities.
5. Teachers reported high levels of engagement with the programme, not only from the pupils and themselves, but also from the senior leadership team (SLT) at their schools. This said, implementing PoP was perceived to be very time intensive.

EEF security rating

These findings have a moderate to high security rating. This was an efficacy trial, which tested whether the intervention worked under developer-led conditions. The trial was a well-designed two-armed randomised controlled trial. However, whilst pupils had similar prior attainment at the beginning of the trial, due to attrition in the original sample some limitations were identified in the final analysis sample, such as the possibility that the control group consisted of higher prior KS1 scores and schools with more positive Ofsted ratings, compared to schools in the intervention group. The trial lost two padlocks (see Appendix B) due to missing data and the number of pupils and schools who started the trial not included in the final analysis. Whilst the trial was well powered, 27.3% of pupils who started the trial were not included in the final analysis because schools declined to participate in testing at the end of the intervention, or because pupils were absent at the point of testing.

Additional findings

Pupils in Power of Pictures (PoP) schools made, on average, one additional month's progress compared to those in the control group equivalent. This is our best estimate, which has a moderate security rating. As with any study, there is always some uncertainty around the result: the possible impact of this programme also includes small negative effects of one month's less progress, and positive effects of up to three months' progress. The programme was also found to have small to medium positive impacts on pupils' writing self-efficacy and idea generation (ideation) scores. Although there was a range of possible impacts in these measurements, the combined positive differences detected indicate that PoP may have had a positive impact on these secondary outcome measures.

Senior leader buy-in and allowing enough time for the programme to be completed were identified as key factors for successful implementation of PoP. In some schools, senior leaders were actively involved with the programme, creating opportunities for teachers to share learning about the programme with other teachers, or attending PoP lessons and training sessions themselves. Despite benefitting from having the clear goal of producing a picturebook, some teachers were unable to complete the programme in the planned time. Some teachers also recommended that guidance be provided to schools and senior leaders on how teachers can deliver the intervention whilst also meeting national curriculum requirements (e.g., spelling, punctuation and grammar).

Teachers reported increased interest in writing, especially from pupils who were struggling to engage with the traditional writing curriculum and may not previously have enjoyed writing or literacy learning or exhibited confidence in their ability as writers. The original PoP logic model suggested that the programme's impact on writing outcomes may be mediated by pupils' engagement with and motivation for writing, which may in turn influence pupils' self-efficacy as a writer. This was well supported by the IPE and is somewhat confirmed by the impact evaluation. Building on the IPE analysis, a new outcome was added to the logic model to represent the greater level of interest pupils developed in picturebooks.

The PoP programme was previously evaluated using a process evaluation with a report released in 2017. The results of that evaluation were highly positive, finding that the 'courses [were] a very impressive success' (Horner & White, 2017). The report specifically highlighted the intervention's role in using creativity to prepare for and develop writing. The findings from this report show suggestive evidence of the programme's positive effect on writing skills, writing self-efficacy and ideation, supported by quantitative evidence not explored within the 2017 evaluation.

Cost

The average cost of CLPE's PoP programme for one school was around £352 for the year it was delivered, or £5 per pupil per year when averaged over three years. This cost estimate is based on the delivery of the intervention to 25 Year 5 pupils per academic year, as this was the average number of pupils treated during this trial. The main costs were due to programme fees, the purchasing of books and materials, and expenses for transporting teachers and pupils to attend training and the workshop, the majority of which are realised in the first year of delivery.

Impact

Table 1: Summary of impact on primary outcome(s)

Outcome/ Group	Effect size (95% confidence interval)	Estimated months' progress	EEF security rating	No of pupils	*P*-value	EEF cost rating
Writing score (all pupils)	0.09 (–0.07; 0.24)	+1	🔒🔒🔒🔓🔓	1945	0.27	**£** £ £ £ £
Writing score (FSM pupils)	0.06 (–0.20, 0.32)	+1	N/A	542	0.67	**£** £ £ £ £

Introduction

Background

This evaluation is part of a round of funding between the Education Endowment Foundation (EEF) and the Royal Society of Arts to test the impact of different arts-based learning strategies in English schools, entitled 'Learning about Culture'.[1] The aim is to improve the evidence base around arts-based education programmes. Learning about Culture consists of five programmes: two in Key Stage 1 (Reception and Year 1) and three in Key Stage 2 (Year 5). Despite the unique aspects of these intervention models, there are many similarities in how they are delivered and what they hope to achieve.[2] The programmes have been supported by Arts Council England.

A focus on increasing attainment in literacy and numeracy has been criticised for leading to a marginalisation of art, music and cultural studies in English schools (Neelands et al., 2015). The UK Government's Culture and Sport Evidence Review (Newman et al., 2010), which summarised much of the observational and qualitative research in this area, showed student participation in cultural learning programmes (from piano training to theatre-based drama projects) to be correlated with higher levels of achievement in mathematics and literacy/English in both primary and secondary school. The review also linked participation in cultural learning programmes to faster language development in the early years and improved cognitive ability. Additionally, large cohort observational studies in the US have suggested that the mathematics and literacy gains associated with cultural participation are particularly large for students from low-income groups (Catterall, 2009, 2012).

Power of Pictures (PoP) developed out of the Power of Reading research (part of the Centre for Literacy in Primary Education, CLPE) and a broad set of literature that established the significance of picturebooks in the development of literacy (Doonan, 1993; Kiefer, 1995; Nikolajeva & Scott, 2013). This was shown to be especially true for pupils with weak literacy skills. The original PoP programme was created with funding from Arts Council England under their Grants for the Arts programme and the support of author-illustrator Ed Vere.

The PoP programme has been evaluated with a report released in 2017. This process evaluation included course observation, surveys of participating teachers, a review of materials, and interviews with author-illustrators. The evaluation was highly favourable of PoP, concluding that the 'courses [were] a very impressive success' (Horner & White, 2017). The evaluators found the programme to have a lasting effect on pupils as perceived by their teachers, and as well as on the author-illustrators who participated. This study did not include a comparison group nor any measure of pupil attainment.

The key findings of their evaluation were:

- discussion before writing helps children to have something to say and the vocabulary to say it;
- giving children the time to draw or express themselves creatively helps them to prepare for and develop their writing;
- teachers understood much more about the writing process having heard authors describe how they developed their ideas;
- teachers found that when they understood the construction of picturebooks they realised how much of the story – including information about character and structure – was conveyed in the pictures;
- picturebooks are far more than just for the young or less able readers;
- authors learnt about the potential of exploring their books in detail with children in the classroom, and how the texts offer 'many rich possibilities for discussion of themes, characters, structure and plot' (Horner & White, 2017).

Building on this limited prior evidence base, this evaluation was designed to estimate the effect of participating in PoP over the course of one school year on pupils' writing skills. This trial was designed as a two-armed clustered randomised

[1] See the RSA website for more information (https://www.thersa.org/globalassets/pdfs/reports/rsa-learning-about-culture-report.pdf).

[2] For an overarching flow diagram of the programme similarities, please see Appendix 1 in the evaluation protocol (https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Protocols/PoP_Evaluation_Protocol_(amended).pdf).

trial with randomisation occurring at the level of the school. This level of randomisation was selected since entire classes participate in the programme and thus the risk of contamination within schools is very high. The two arms were (a) participation in the PoP (Treatment), and (b) business as usual (Control). Blocking was used in the randomisation to improve cross-arm comparability of schools, to improve precision of estimates, and to allow schools that sign up early to receive their allocation sooner than they otherwise would. More detail on the intervention is provided in the *Methods* section of this report. Initially, we had planned to look at the long-term effects of participating in the PoP after one further year, looking at results from the end of Key Stage 2 national curriculum tests in English grammar, punctuation and spelling.

We note upfront that it has been necessary for the analysis of this trial to deviate substantially from our initial plans set out in the project protocol and statistical analysis plan (SAP). These stem from issues in accessing the pre-test data that we expected to be able to obtain from the Department for Education (DfE)'s National Pupil Database (NPD). During the implementation of the trial, the DfE changed the way in which data from the NPD are made available to researchers, switching from providing extracts that can be used alongside project data within evaluator's own secure computing systems to requiring access within the Office for National Statistics (ONS) Secure Research Service (SRS). In turn, this means that it is now necessary for project data to be uploaded to the SRS. Given that this project data is considered personal data over which we as evaluators are data controllers, this requires the conclusion of an appropriate data sharing or processing agreement between the evaluator and the DfE and/or the ONS in order to provide legally required reassurance by the DfE/ONS about the treatment of personal data over which the evaluator is controller.

Due to extended negotiations and delays between the evaluators and DfE/ONS, which we understand to have been severely exacerbated by additional workload due to the COVID-19 pandemic, in the interests of completing these evaluations and after discussion with the EEF and project teams, the decision was made to proceed with analysis with deviations from protocol flagged as we move through the methods section. These deviations were agreed with the EEF ahead of conducting the analysis. Beyond issues inherent in deviating from pre-registered protocol, the main implications for the analysis are reductions in the statistical power relative to expectations.

It is important to understand the implications of this change. The purpose of including baseline measures in the current evaluation is to increase its statistical precision (i.e., to reduce the uncertainty around intervention impact estimates), which makes them more likely to be statistically significant. Importantly, both the original and the substituted baseline measures are taken from prior to the randomisation and intervention. Therefore, due to the randomised nature of the evaluation, their inclusion does not bias any intervention impact estimates, but only affects the statistical uncertainty around these estimates (i.e., the extent to which they are detectable as statistically significant). Given EEF policy to report impact estimates whether or not they are statistically significant, there is an increased risk that headline positive or negative effects are just due to this uncertainty, rather than representing a true effect. As a result, we particularly stress the importance of statistical significance as a check on interpretation of the results in this report.

Intervention

1. **Brief name.** Power of Pictures (PoP)
2. **Why (rationale/theory).** The PoP intervention aims to develop primary school pupils' writing skills by training teachers to use picturebooks in their lessons. Through its programme, PoP aims to enhance teachers' understanding of the power of picturebooks and increase teachers' comfort in, and ability to teach, using picturebooks to a range of ages (see Figure 1 for a logic model of the intervention). The intervention includes 2.5 days of training for teachers directly from the author-illustrator and the PoP team. The improvement in teachers' skills is thought to have an effect on pupils' writing skills.
3. **What (materials).** Resources and materials[3] for the intervention included (see Figure 1, *Activities*):
    - *Picturebook* from selected author-illustrator;
    - Web resources available to the teachers, including sample teaching sequences (available at **https://clpe.org.uk/powerofpictures**).
4. **What (procedures).** The PoP team has a group of picturebook authors with whom they work. Their selection is based on the quality and relevance of a specific book from the author's work which will be used for the

[3] See https://clpe.org.uk/powerofpictures for all resources relating to the Power of Pictures (PoP) intervention.

intervention. The PoP team works with the author-illustrator (always the same person) of this book throughout the course of the intervention. The author-illustrator is selected based on the quality of their text and their experience of talking about their process, either through direct teaching, workshops at festivals or similar. They take part in training with the PoP team before the intervention in the schools begins.

Schools that are selected to take part in the intervention must commit at least four weeks of their English curriculum (20 lessons) to PoP. One teacher in Year 5 is nominated by the school to lead the intervention for their class.

The teachers selected from each school receive 2.5 days of training out of school time between October and March directly from the author-illustrator and the PoP team. The first day (in October) of training focuses on conveying meaning through drawing, reading pictures, creative approaches to using books and feedback. Throughout the entire process, the PoP team makes web resources available to the teachers, including sample teaching sequences.

Following the first day of training, the teachers are expected to do a 'gap task' with their pupils, where the pupils work with a picturebook by the focus author-illustrator. This task is documented via photos, which are sent to the PoP team. During this time (in November), the participating class attends a half day visit delivered by the PoP team and the author-illustrator, where they get to meet the author-illustrator and take part in a workshop focused on character design and development of narrative. The visit gives the pupils the opportunity to meet the author-illustrator.

The second day of training for the teachers, which occurs in January, focuses on the full writing process for a book, how to design characters, sequential storytelling, feedback on writing, publication and reflection. This training session is delivered by the same PoP team and author-illustrator as the first training session. The teachers then return to their class for another six-week period and use the teaching sequences in more detail to focus on the development of pupils' own picturebook stories, inspired by the workshop. This is again documented via photos sent to the PoP team. These materials were not used as part of the evaluation.

The third and final training session (lasting half a day) involves not only the teachers who have previously participated, but also the school's senior leadership team (SLT) and the literacy coordinator. This session takes place close to the end of the intervention. In this session, the school staff reflect on the effectiveness of the PoP approach and an action plan is formulated for how to continue incorporating picturebooks and illustration into schools' curriculum, based on successes shared.

5. **Who (recipients).** PoP is targeted at Key Stage 2 (Year 5) pupils.
6. **Who (implementers).** PoP is delivered by the Centre for Literacy in Primary Education (CLPE). The programme developed out of the Power of Reading research. The original PoP programme was created with funding from the Arts Council England under their Grants for the Arts programme and the support of author-illustrator Ed Vere.

The CLPE team is led by Charlotte Hacking. Charlotte is an experienced teacher and senior leader. She previously taught across the primary school age range before holding several posts including Assistant Headteacher and leadership posts in Literacy, Early Years (EY) Foundation Stage, more able pupils and KS2. She was a literacy consultant within a local authority, focusing on EY, phonics and primary literacy. In 2020, Charlotte joined the Helen Hamlyn Centre of Pedagogy as a member of the Practitioner Advisory Board. She developed and leads PoP research at CLPE, and also led and developed CLPE's Power of Poetry research project.

The author-illustrators are all published picturebook authors who are selected by CLPE on the basis of their work. The three author-illustrators who participated in this trial were Viviane Schwarz, Alexis Deacon and Tim Hopgood. Their training is based on a planned and tested model of training, which received Arts Council funding. The project has been developed and delivered over the course of six years with 10 author-illustrators. Details of the writing process model can be found here:

https://clpe.org.uk/powerofpictures/teaching-approaches/following-authentic-writing-process

The participating teachers were selected by the treatment schools. One Year 5 teacher was selected per school. They receive training from CLPE before implementing the PoP intervention in their classes.

7. **How (mode of delivery).** Training was delivered face-to-face by the PoP team to the participating teachers. Sessions were delivered face-to-face by the Year 5 teacher to his/her entire class.
8. **Where (setting).** Sessions were delivered in the Year 5 classroom. The half day workshop for the pupils to meet the author-illustrator took place in a classroom in specifically chosen arts/literature-based venues in each location. The training sessions for the teachers also took place at CLPE in London and Seven Stories in Newcastle.
9. **When and how much (dosage).** PoP activities were incorporated into English lessons over the course of one school year. Schools that participate must commit at least four weeks (20 lessons) of English instruction time to the intervention.

   There is a total of 2.5 days of continuous professional development (CPD) training for participating teachers. Each of the two full CPD days ran from 9.30am–3.30pm and from 9.30am–12.30pm on the final half day.

   The intervention has two specific two-week tasks, designed to be completed in the six-week gap between teacher training sessions ('gap tasks'). The first gap task (block one) was documented by the teacher via photos sent to the PoP team (see Figure 1, *Activities*). The second gap task (block two) was again documented by the teacher via photos sent to the PoP team and by teachers bringing examples of the children's completed picturebooks to the final training session. The teachers have some discretion as to how to implement the gap tasks, but rely on teaching sequences for each block provided by the PoP team.

   The children's author workshops ran for 1.5 hours at either 10.30–11.45am or 1.15 – 2.30pm.

10. **Tailoring.** The intervention was not planned to be personalised, meaning that all teachers received the same intervention.
11. **Modifications.** No modifications were made to this intervention.
12. **How well (planned).** According to the PoP model, delivery with fidelity involved teachers delivering ten PoP lessons during the autumn and spring terms (meaning a total of 20 lessons).

This evaluation examines two implementation dimensions: compliance and implementation fidelity.

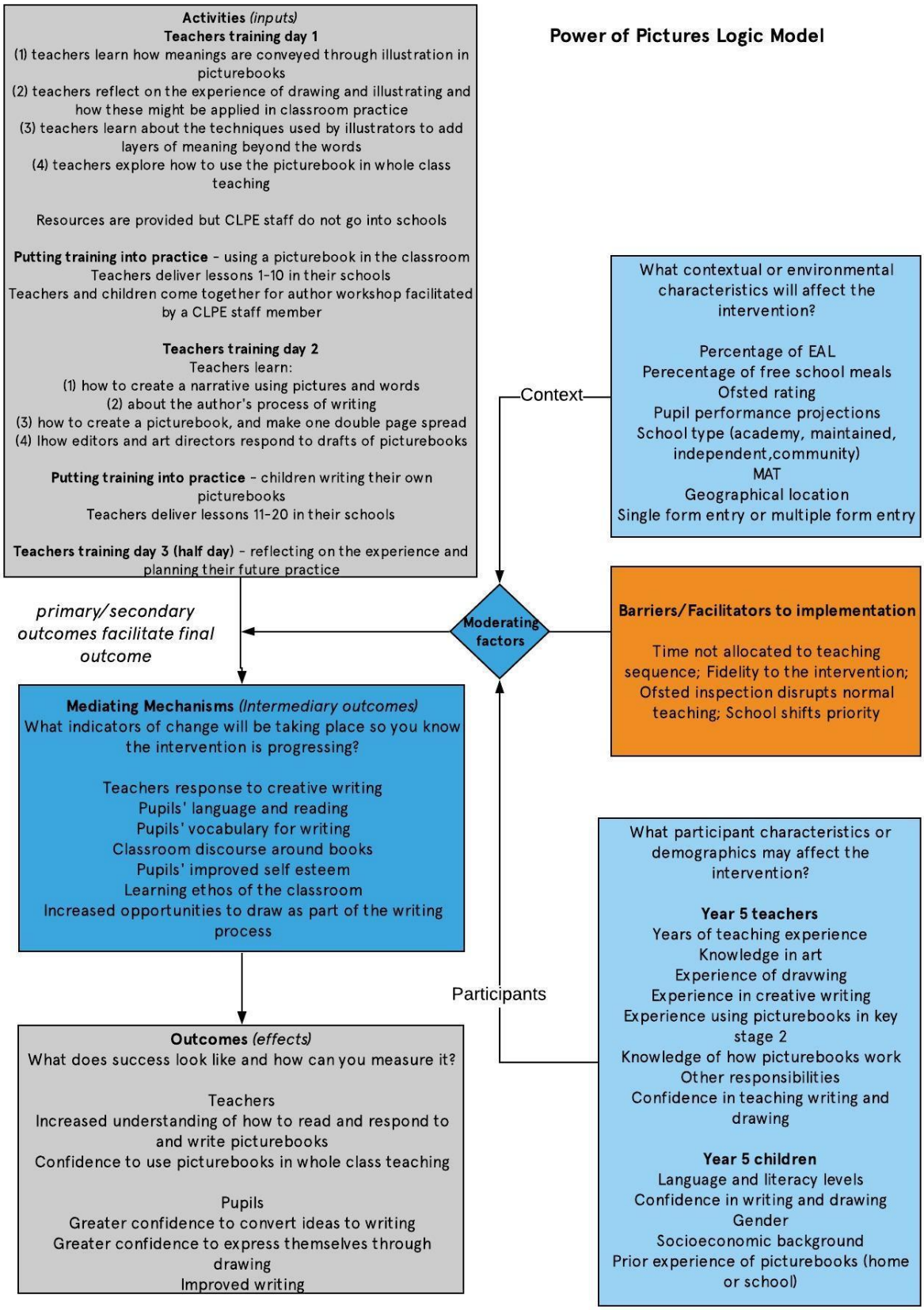
Specifically, compliance with this intervention was measured at the teacher level, which reflects the intervention delivery method. A school is considered to have complied if and only if the following three conditions are met:

1. the participating teacher attended all training sessions (2.5 days);
2. the participating teacher delivered at least 50% of lessons in both block one and two (10/20);
3. the participating teacher's class attended the half day author-illustrator off-site workshop.

Compliance data was collected by CLPE.

In comparison, implementation fidelity was conceptualised as how the way in which the intervention was implemented in practice compares to the intended implementation of the intervention as described in this section of this report.

Figure 1: Logic model (original version)



Evaluation objectives

The primary objective of this evaluation is to estimate the effect of participating in the PoP over the course of one school year on pupils' writing skills.

The evaluation aims to answer the following questions:

What is the effect of participating in the PoP over the course of one school year on pupils' writing skills? [primary research question]

1. What is the effect of participating in the PoP over the course of one school year on pupils' writing self-efficacy? [secondary research question]
2. Does participating in the PoP over the course of one school year have an impact on pupils' perception of their own capacity to generate ideas? [secondary research question]

These research questions were reported in the evaluation protocol (initially published in May 2018 with revisions for clarity published in April 2019, **https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Protocols/PoP_Evaluation_Protocol_(amended).pdf**) and further details on the quantitative approach were provided in a SAP (published in December 2018, **https://educationendowmentfoundation.org.uk/public/files/Projects/PoP_SAP_(final).pdf**).

The IPE was designed to explore overarching implementation questions across all five Learning about Culture trials, as well as RQs specific to PoP. The four overarching questions were written based on cross-project similarities; however, not all questions apply to each programme, due to variations in training and delivery (see Appendix Q for the diagram depicting cross-project similarities). The overarching IPE research questions are as follows (question 4 is not applicable to PoP: detailed below).

1. In what ways was the programme implemented? What are the barriers and facilitators of delivery (Fidelity)? In particular:
    - **a)** Senior Leadership Team (SLT) buy-in;
    - **b)** Delivery of training: **(i)** the extent to which it is consistent across sites; and **(ii)** whether it appears to be effective in ensuring that teachers understand the aims and main features of the intervention;
    - **c)** Delivery of the intervention: **(i)** whether it is consistent across sites; **(ii)** whether it appears to be effective in supporting children's attainment; **(iii)** whether it appears to facilitate children's engagement.
2. To what extent did the schools engage with the intervention in line with the intervention aims? (Responsiveness).
3. How was the quality of the intervention perceived by teachers, senior leaders and teaching assistants? (Quality)
4. *[Not applicable]* To what extent is the knowledge of arts practitioners delivering the intervention integrated with the pedagogic knowledge of teachers involved? (Implementer support system). This question is not applicable because PoP does not involve direct delivery in schools by arts practitioners.

In addition, the IPE also sought to answer questions specific to the PoP intervention. Where there were deviations from the protocol, they are noted in relation to the question and described further below.

5. How does a pupil's home learning environment and exposure to books affect their engagement? (Characteristics)
6. How does teachers' response to creative writing change as a result of taking part in the training? How is this reflected in teacher confidence in writing and drawing, and in classroom practice in PoP lessons and more generally?
7. How did teachers experience and perceive the quality of the PoP training? *[changed from protocol]*[4]

[4] Question from protocol: In what ways do the illustrators' delivery of the PoP training vary? How does this variation impact on teachers delivering PoP?

8. How did pupils engage with the author-illustrator workshop element of the programme? *[changed from protocol]*[5]
9. Are there any unforeseen mechanisms in the intervention that appear to be bringing about change? (Mechanisms)
10. What assumptions about the quality of texts are built into the intervention, and how is this reflected in teacher views and practice (e.g., some authors and types of texts regarded as better quality than others)? (Quality) *[not answerable]*

Questions 7 and 8 were changed from the protocol due to data availability and improved understanding of the programme as the fieldwork progressed. For question 7, we did not have sufficient data to compare illustrators' delivery of training, so instead focused on teachers' reports of their experience and perceptions of training. For question 8, this question was broadened as a result of the data and analyses. The pupils' experience of attending and engaging with the workshop came through as of more importance than the specific author-illustrator who ran the workshop. Question 10 was omitted at reporting stage due to lack of data. Interview data from teachers did not produce detailed insights about the quality of the texts.

We had also planned to estimate the longer term effects of participating in the PoP after one further year, looking at results from the end of Key Stage 2 national curriculum tests in English grammar, punctuation and spelling. However, it will not be possible to carry out this analysis due to the cancellation of these assessments as a result of the COVID-19 pandemic.

Ethics and trial registration

The project's aims, methods and materials were reviewed through the processes laid out by the UCL Institute of Education research ethics committee and approved on 14 December 2017. Although the application was approved, the ethics reviewers stressed the importance of ensuring ongoing pupil assent for participation in any evaluation activities throughout the research. As such, all research assistants (RAs) conducting assessments with pupils verbally described the activities to the pupils using age-appropriate language, informed them all activities were voluntary, and gave an opportunity for pupils to decline to participate.

Schools were informed about the trial through initial information from the developer and formally committed to participation by signing a memorandum of understanding (MoU). (A template version of this document is included as Appendix M to this report.)

This trial protocol has been pre-registered at www.controlled-trials.com, and assigned an International Standard Randomised Controlled Trial Number (ISRCTN) of ISRCTN15334278.

Data protection

As part of this project, we processed pupils' and teachers' personal data. For this reason, it was important that we processed this data lawfully, following the principles laid out in the Data Protection Act 1998 (DPA) until May 2018 and the EU General Data Protection Regulation (GDPR) from May 2018 until December 2020, and the UK General Data Protection Regulation from January 2021 (the project spanned these three regulatory periods). We explain the lawful basis below with respect to the GDPR, but there are equivalent regulations in the DPA for the justifications set out below.

BIT used Article 6(1)f of the GDPR as the lawful basis for processing personal data as part of this project. This is generally known as the 'legitimate interests' basis. There's a guide to that article here: **https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr*/lawful-basis-for-processing/legitimate-interests/**. BIT carried out a 'legitimate interests assessment' in support of this, identifying societal benefits to this

5 Question from protocol: Does pupils' engagement differ with different author-illustrators? Does pupils' engagement have any relation to their self-esteem and confidence in relation to writing?

processing of personal data. Specifically, the use of pupils' and teachers' personal data as part of this research was to understand the benefits to pupils of participating in this programme in terms of their academic attainment and other related benefits. This has public benefits that BIT believes are significant in terms of understanding whether this programme has the potential to benefit children in schools across England. Without processing these data it would not have been possible to provide this quality of new evidence.

UCL used Article 6(1)e of the GDPR as the lawful basis for processing personal data as part of this project. This is generally known as the 'public task' basis. UCL has reviewed current ICO guidance available here: **https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/public-task/**, and has determined that this research forms part of its performance of a task in the public interest, as one of its core purposes provided for in its Charter and Statutes. This use of data was allocated the following UCL Data Protection Registration Number: Z6364106/2017/11/69 social research.

We do not believe that any of the data we processed falls within the definition of special category data under the GDPR. This would have required an additional justification under Article 9(2) of the GDPR.

We informed pupils' parents of the proposed data processing and provided the opportunity fo them to object to this. If parents objected, then the pupils' data was never passed to us by schools. If a parent chose to withdraw their child's data at a later stage then it was destroyed. The data controllers were named in the privacy information provided as part of this project and contact details provided in case they had any queries about the data we hold about them, including provision and deletion of their data. The relevant letters and forms have been reproduced in Appendix P.

The information provided to parents explained in clear and plain language the lawful bases for processing, the purpose to which we put the data, that they could object to this data processing and this would be respected, contact details of the organisation, and categories of data that have been processed.

Data will be kept until the end of the research project, including academic paper writing and dissemination (and certainly not longer than 10 years, in line with UCL's policy on data retention). When it is deleted, it will be securely destroyed.

Data will be shared with the Education Endowment Foundation (who funded the trial), EEF's data contractor FFT Education (who manage the EEF's Data Archive) and (in a form that is truly anonymised) to the UK Data Archive. Details of this sharing were included in relevant data privacy notices.

Project team

The project team comprised Charlotte Hacking and the Primary Advisory Teachers (Darren Matthews, Katie Myles and Jonathan Rodgers) at the Centre for Literacy in Primary Education (CLPE). The intervention was designed and delivered by the CLPE with support from their author-illustrator partners.

The impact evaluation was led by Dr Jake Anders and Dr Nikki Shure at UCL Institute of Education with support from Prof. John Jerrim, and was led at the Behavioural Insights Team by Kimberly Bohling and Dr Matthew Barnard. Data collection was managed by Faisa Abdi, Eleanor Collerton, Camilla Devereux, Amber Evans, Louise Jones, Alex Manby, Bridie Murphy and Juliane Wiese of BIT. Primary data collection was carried out by RAs employed by BIT and marking of those data was also carried out by RAs employed by BIT, drawn from finishing students at UCL Institute of Education. The IPE was led by Dr Matthew Barnard at the Behavioural Insights Team with support from Faisa Abdi and Johanna Frerichs and input from Prof. Dominic Wyse (UCL IPE lead), Prof. Gemma Moss and Prof. Andrew Burn at UCL Institute of Education. The cost evaluation was led by Dr Matthew Barnard with support from Faisa Abdi of BIT. The evaluation design was also supported by Daniel Carr, Dr Florentyna Farghly, Dr Jessica Heal and Dr Pantelis Solomon of BIT.

Methods

Trial design

Table 2: Trial design

Trial design, including number of arms		Cluster randomised, two arms
Unit of randomisation		School
Stratification variable(s) (if applicable)		Proportion of English as an additional language (EAL) students (high/low split at sample median); proportion of free school meals (FSM) students (high/low split at sample median)
Primary outcome	Variable	Writing attainment
	Measure (instrument, scale, source)	Writing assessment measure (WAM), 0–32, (Dunsmuir et al., 2015)
Secondary outcome(s)	Variable(s)	Writing self-efficacy Ideation
	Measure(s) (instrument, scale, source)	Writing self-efficacy measure (WSEM), 16–80, (adapted from Bruning et al., 2013) Ideation captured from sub-measure of the WSEM, 5–25.
Baseline for primary outcome	Variable	Planned to be:[6] Phonics attainment Protocol deviation: FSM status, EAL status, class FSM composition, class EAL composition
	Measure (instrument, scale, source)	Planned to be: Phonics Screening Check (DfE) Protocol deviation: 0/1 indicator of FSM eligibility status, 0/1 indicator of EAL status, proportion of class reported FSM eligible [0,1], proportion of class reported EAL [0,1]. (All derived from school reports collected ahead of randomisation; see p.18 for further details and justification).
Baseline for secondary outcome(s)	Variable	Planned to be: Personal, social and emotional development skills Protocol deviation: FSM status, EAL status, class FSM composition, class EAL composition

[6] See *Background* section.

	Measure (instrument, scale, source)	Planned to be: Early Years Foundation Stage Profile personal, social and emotional development skills (DfE) Protocol deviation: 0/1 indicator of current FSM eligibility status, 0/1 indicator of EAL status, proportion of class reported FSM eligible [0,1], proportion of class reported EAL [0,1]. (All derived from school reports collected ahead of randomisation; see p.18 for further details and justification).

This trial was designed as a two-armed stratified, clustered randomised efficacy trial with randomisation occurring at the level of the school and outcomes measured at the level of pupils. This level of randomisation was selected since entire classes participate in the programme and thus the risk of contamination within schools is very high. The two arms are as follows:

- Participation in the PoP (Treatment)
- Business as usual (Control)

The primary outcome of interest is improvement in pupils' writing, with the secondary outcomes of their writing self-efficacy and writing creativity (Ideation). Further information on the approach taken to measuring these outcomes is provided below.

Participant selection

The project aimed to recruit 120 English state-funded primary schools from the following geographical regions: London, North East England (training located in Newcastle), and East Sussex (training located in Brighton). These locations were agreed with the developers during initial setup meetings due to the locations of their author-illustrators and existing networks. More than 120 schools were successfully approached by CLPE; however, due to GDPR-related concerns among some key local authorities, and other schools finding it challenging to submit pupil data within the timeframe needed, this substantially reduced the number able to be fully recruited. In the end, 101 schools were fully recruited and included in the randomisation.

PoP is currently delivered to pupils across the primary school age range. Year 5 was chosen for the purposes of the evaluation, given the greater perceived potential to capture writing-based outcomes from pupils at older ages, but without attempting to deliver and evaluate in Year 6 classes because of the perceived tension with Key Stage 2 national curriculum tests at the end of the year. Furthermore, evaluation of Year 5 delivery was intended to allow for medium-term follow up in those Key Stage 2 national curriculum tests, but this was ultimately impossible due to their cancellation for the relevant year group in the wake of COVID-19.

One teacher of a Year 5 class in each participating school was nominated for participation in the trial. All children in the teacher's class participated in the trial (other than where objections to processing personal data were received or lack of consent to participate in evaluation activities). Except in unforeseen and unavoidable circumstances (e.g., teacher moving school), the teacher (and, therefore pupils) who participated were selected prior to randomisation, to minimise the potential for this to introduce differences between the intervention and control groups; except in a small number of cases (two schools), all data on participating pupils was collected pre-randomisation in order to ensure this.

In order to be considered for participation in the trial, schools had to agree to: (i) distribute opt-out consent[7] forms to parents; (ii) provide pupil data; allow for linking between trial data and the DfE's NPD (for pre-test data and long-term follow up); (iii) identify the teacher who would participate in the trial; and (iv) cooperate with the project and evaluation

[7] Note that this is opt-out consent from a research ethics point of view, not from a data protection point of view. We note that since the first version of this protocol was agreed the GDPR has been implemented. As such, UCL's legal basis for processing this data is now considered to be 'public task' and BIT's legal basis for processing personal data is now considered to be 'legitimate interest'. 'Consent' is not used by either party as a basis for the processing of personal data. Nevertheless, it remains the relevant term in respect of research ethics.

teams during the trial (further details of these requirements are outlined in the MoU with Schools, available in Appendix M).

The PoP team advertised the trial on its website and also approached schools through their existing networks. The team aimed to recruit schools that had larger populations of individuals receiving free school meals (FSM) than the national average of 15.3 percent of pupils aged 5–10 (DfE, 2016). This appears to have been successful, judging by the proportion of pupils eligible for Free Schools Meals in the analysis sample.

The eligibility criteria for schools to participate were:

- participating schools must be English state-funded primary schools recruited from the following regions: London, North East England and East Sussex;
- schools had to agree for one Year 5 class to participate in the intervention and commit at least four weeks of their English curriculum to the intervention;
- schools had to nominate one Year 5 teacher to lead the intervention, who would also attend the PoP training sessions;
- schools had to agree to distribute study information sheets, data privacy information, and data processing objection forms to parents;
- schools had to agree that, if allocated to the control group, they would continue with 'business as usual' for the duration of the trial;
- schools had to return a signed MoU, and commit to participate fully in the study – including the collection of outcome measures in summer 2019 – regardless of which trial arm they were assigned to;
- schools had to agree to allow time for each assessment phase and liaise with the evaluation team to find appropriate dates and times for assessments to take place;
- schools had to agree that teachers in both trial arms cooperate with activities for the IPE, if requested.

Outcome measures

Baseline measures

Baseline measures for this research were planned to be drawn from the DfE's NPD. All participating schools were asked to provide personal information about pupils in the participating teacher's class which would allow a reliable link to be achieved, based on current guidance from the DfE and balancing this against personal data minimisation requirements set out in data protection legislation. Using this link it was expected that we would obtain information on pupils' performance in the phonics screening check (using the NPD variable PHONICS_MARK) for the primary outcome analysis, and assessments of pupils' personal social and emotional development from the Early Years Foundation Stage Profile (aggregated scores from NPD variables FSP_PSE_G06, FSP_PSE_G07 and FSP_PSE_G08) for the secondary outcome analyses.

Due to the data access issues described above, an alternative approach was taken, with its design informed by an intention to maximise the explanatory power of our analysis model and, hence, the precision of our treatment estimates, given the data available. As such, instead of including the planned baseline measures in the model, we substituted the available demographic information that was collected about pupils ahead of randomisation (initially intended for the purposes of stratification/blocking as part of the randomisation process), specifically eligibility for FSM and whether the child has English as an additional language (EAL). We included these in the model as predictors, and also aggregated them to the class level to produce composition variables, given evidence that school-level aggregate predictors also provide explanatory power (Bloom et al., 2007). FSM and EAL status are both known to predict academic attainment (Strand et al., 2015; Sutherland et al., 2015) and, as such, we expected this to improve power compared to an empty model.

Nevertheless, the improvement in statistical power is still likely to have fallen short of that we would have expected from including a prior attainment measure, as was planned. It is important to understand the implications of this change, and especially to stress that there are no expected implications for bias in our impact estimates from not having our planned baseline measures – the unbiasedness of RCT estimates derives from the randomisation, not from statistically

controlling for differences at baseline. Indeed, in principle, there is no need to include any baseline measures at all in the analysis to achieve an unbiased estimate from an RCT. Inclusion of inappropriate covariates in our analysis would have the potential to introduce bias – such inappropriate covariates are ones that could have been affected by the treatment, which is why we are including pupil characteristics from prior to randomisation. The main implication of this change is a reduction in statistical precision (i.e., the uncertainty around estimates that is inherent in all evaluations is likely to be larger in this evaluation than it would have been), which is manifested as wider confidence intervals (or, equivalently, less likely to be statistically significantly different from zero for a given size of impact estimate). Given EEF policy to report impact estimates whether or not they are statistically significant, there is an increased risk that headline positive or negative effects are just due to this uncertainty rather than representing a true effect that would have been the case in the presence of more explanatory power from baseline measures. As a result, we particularly stress the importance of statistical significance as a check on interpretation of the results in this report.

Primary outcome

Writing attainment

To measure the primary outcome, we used the writing assessment measure (WAM) (Dunsmuir et al., 2015; Murphy et al., 2013). The WAM was developed in order to create a valid and reliable writing assessment tool, relevant within the context of the English educational system. This measure is designed to assess narrative writing in response to a written prompt. When the measure is administered, pupils are given 15 minutes to write their response. Previous evidence suggests that this measure is reliable (test–retest correlation $r = 0.82$ over 21 days with different prompts) and valid (concurrent validity correlation of $r = 0.786$ with Wechsler Objective Reading Dimensions (WOLD) Written Expression subtest) (Dunsmuir et al., 2015). The WAM prompt presented to pupils is included as Appendix N. Given the WAM's limited use up to this point, we carried out further analysis of the performance of the measure, the results of which are reported below.

The WAM is based on the structure and format of the WOLD Written Expression subtest, with modified dimensions that incorporate descriptors from the national curriculum writing attainment targets, including the following seven dimensions: ideas development; organisation and planning; vocabulary; sentence structure; grammar; spelling, punctuation and handwriting. For each of these sub-scales, the pupil can receive a mark of 1–4, with 4 being the highest. The WAM is unique as an assessment because it incorporates 'ideas development', which fits well with aims of the interventions being tested as part of the Learning about Culture project. Note in the logic model for this intervention (Figure 1) that increased creativity is an outcome. In support of this, we double-weighted the score on the 'ideas development' dimension as part of the impact analysis. Final scores range from zero to 32 (after accounting for double-weighting).

Robust assessment of writing is challenging, particularly during primary schooling. However, the centrality of understanding improvements in writing and, hence, the need for this to be the primary outcome measure was stressed in the initial project outlines (noting that previous trials had generally focused on reading, rather than writing, adding to the rationale for funding these projects) and setup meeting discussions with the EEF and programme teams. Use of the WAM (Dunsmuir et al., 2015) as a measure of KS2 pupils' writing was not the initial proposal for this trial but emerged from discussions held as part of the project setup meetings. The WAM is an analytic measure of writing based on equal weighting of the following criteria: handwriting, spelling, punctuation, sentence structure and grammar, vocabulary, organisation and overall structure, and ideas. There are, of course, some limitations of its use, largely stemming from the fact that it is a relatively new measure and we would, ideally, prefer to have used a measure with a longer track record. Nevertheless, we believe that it is a pragmatic measure for the context of this research. Dunsmuir et al.'s (2015) results are encouraging in terms of the measure's internal consistency (Cronbach's alpha = 0.87), inter-rater reliability (Cohen's kappa > 0.7 for all sub-scales except 'ideas', where kappa = 0.62; kappa > 0.6 is generally considered satisfactory), and test–retest reliability ($r = 0.82$). To supplement this evidence, the EEF provided funding for us to undertake a small, informal piloting of the WAM.

The Behavioural Insights Team (BIT) conducted a small-scale pilot of the WAM in October 2017 with approximately 50 pupils from one Year 5 class, one Year 6 class, and one mixed Year 5/6 class. The aim was to understand how clearly pupils understood the prompt, how much they were able to write during the time allotted, and act as a sense check of the measurement properties reported by Dunsmuir et al. (2015). Pupils were given the WAM prompt, one sheet of A4 paper, and 15 minutes to complete the task. The results of the pilot showed that pupils had little difficulty in completing

this writing task but required some additional clarification on the prompt and additional paper. In addition, the measurement diagnostics remained encouraging (albeit this may have been helped by the small sample).

While we were keen to maintain consistency between the WAM as implemented by Dunsmuir et al. and this work, in order to ensure that what we do know about the WAM from their work carries forward to our context, we have made some small adaptations based on concerns identified during the setup process and from observations arising from the small-scale pilot. These were discussed with the WAM's lead developer, whom we gratefully acknowledge as having provided helpful informal advice as part of this process. Concerns from developers were noted and discussed with mitigating actions as follows:

- **Concerns that the WAM may over-weight surface features of the language.** We do agree that this is a potential concern, particularly within the context of arts-based learning evaluations, while noting that we think alternatives (such as using KS2 grammar punctuation and spelling tests) would be much worse affected by this problem. In partial mitigation we have double-weighted the ideas sub-scale. This double-weighting results in an outcome distribution that is slightly less normally distributed but not to an excessive degree (see Figure E1 in Appendix E).
- **Concerns regarding the content and face validity of the WAM, given its alignment with an earlier version of the English national curriculum.** The overall aims for the teaching of writing in primary schools, that are specified in England's national curriculum implemented since 2014, require that teachers develop pupils' competence in 'transcription (spelling and handwriting)' and 'composition (articulating ideas and structuring them in speech and writing)' (DfE, 2013, p. 16). The WAM is an appropriate measure of writing in the context of England's current national curriculum aims.
- **Concerns that aspects of the prompt may be confusing to pupils.** These were identified from the pilot. Revisions were made to the introduction to pupils of the WAM to provide increased guidance to pupils on the purpose of the writing sample we ask them to produce, given that we understand this to be normal practice for pupils of this age when taking part in a writing activity (see the WAM prompt used in this trial in Appendix N).
- **Concerns that 15 minutes is not long enough.** The observations from the pilot suggested that 15 minutes was insufficient for pupils to produce a writing sample that could be meaningfully assessed. However, clearly this was only small scale. We take this concern seriously while wishing to maintain broad alignment with previous use of the WAM. As such, in a change from the WAM as previously used, we provided five minutes of planning time at the start of the activity in which pupils can make notes but don't begin the writing activity itself. This also helps to make the activity more familiar to pupils, in line with the previous point.

The writing tasks for this evaluation and the other two Key Stage 2 evaluations (Craft of Writing and Young Journalist Academy) that were part of the EEF/RSA Learning about Culture project were invigilated and collected in summer 2019 by a team of research assistants (RAs) coordinated by BIT as a combined exercise. Since outcome data was collected as part of a single exercise and consistent (in terms of both measure and timing) across these three evaluations we report our analysis of the data collection and measurement with pooled WAM data collected across the three projects; pooling these data allows us to increase sample size for these analyses and, hence, reduce noise and risk of small sample bias in estimates from these analyses. The same goes for consideration of the WSEM measure, which follows below beneath the secondary outcomes subheading.

Data collection RAs were kept blinded to trial arm assignment of schools in order to avoid the potential for this to bias the outcome measurement, for example by being more lenient on timing in treatment schools. A separate group of 25 RAs (20 of whom marked tasks on the PoP project), also blind to trial arm assignment of schools (this time in order to avoid the potential for this to introduce bias into the trial, for example through unconsciously being more generous in their marking of pupils in the treatment group), marked the writing exercises against the WAM scoring sheet. This blinding is important in supporting the evaluation's internal validity. The WAM scoring sheet provided detailed criteria for assigning scores on each of the seven dimensions. During training, markers were provided with examples of student writing that exemplified each rating within a given dimension and an explanation of why that sample achieved the rating. The writing samples, scores and explanations were all provided by IoE academics who have expertise on the WAM. A random sub-sample of the tests (approximately 3%) were independently second marked by one of the other RAs during this process (a minimum of two tests per RA per day), with a correlation of 0.75 between markers in this double-marked sample. Where discrepancies did arise these were used to feed back to markers in order to improve the consistency of marking over the exercise as a whole. This continuous improvement process may inflate this correlation over the

course of the process, but improves the reliability of the marking relative to the alternative. As a point of reference, Dunsmuir et al. (2015) report an average marker-level intra-class correlation of ICC = 0.97 for the WAM (range 0.93–0.99 at 95% confidence interval).

Furthermore, analysis of the basic statistical properties of the overall measure are encouraging: the distribution of the scores is normal (skewness = –0.27; see also histogram in Figure E1 of Appendix E), which suggests there were minimal issues with floor or ceiling effects, and we calculate a Cronbach's alpha across the seven marking sub-domains of 0.84, suggesting these cohere sensibly into the overall score.

Secondary outcomes

Writing self-efficacy

As highlighted in the logic model, the impact of the intervention on writing outcomes may have an effect through pupils' engagement with and motivation for writing, which may in turn have an effect on their sense of efficacy as a writer. For this reason, we consider writing self-efficacy as one of the secondary outcomes. In addition, EEF's review of non-cognitive skills highlights that the evidence 'indicate[s] that self-efficacy for a particular task is malleable and that improved self-efficacy is associated with greater persistence, interest, and performance' (Gutmann & Schoon, 2017, p.11) and that 'the best predictors of specific academic performance are self-efficacy beliefs regarding those specific academic domains' (Pajares, 1996, in Gutmann & Schoon, 2017, p.11).

To measure writing self-efficacy, we used a writing self-efficacy measure (WSEM) which was adapted from the self-efficacy for writing scale (SEWS) measure proposed by Bruning et al. (2013), in order to make it suitable for primary school pupils through some simplification of language (see Appendix O for the full WSEM questionnaire used in this trial). Bruning et al.'s original measure involves 16 statements capturing aspects of writing, including 'I can think of many ideas for my writing' and 'I can avoid distractions while I write', with pupils giving marks out of 100 for their self-assessment in each of these. We used slightly simplified versions of some of these statements to better suit the primary school context; in addition, we requested responses on a five-point Likert scale. rather than marks out of 100 These adaptations were based on consultation with experts in primary literacy pedagogy and were piloted through the same process as outlined for the WAM above, with resulting refinements to wording of the Likert categories as the initial versions were found to encourage pupils to exaggerate their confidence as this was seen as the 'right' answer. Bruning et al. (2013) developed a multi-factor model of writing self-efficacy; however, since the intervention is not hypothesised to have a link with the specific factors (with the exception of ideation, which we discuss separately below) we used a simple aggregate of self-assessments across all sixteen statements (all are positively framed so there is no need for reverse coding). As such, possible scores range from 16–80 for each child.

As with the WAM, these tasks were administered and collected in summer 2019 by a team of RAs coordinated by BIT, who also marked the WSEM. RAs were kept blinded to trial arm assignment of schools. Again, consistent with the WAM, we explored the statistical properties of this measure (carried out pooled with WSEM data collected concurrently for the Craft of Writing and Young Journalist Academy projects), given the adaptations that were made in order to use it for this project. The overall scores are rather negatively skewed (skewness = –1.16; see also histogram in Appendix E) which could attenuate impact estimates for this outcome; we estimate a Cronbach's alpha of 0.90 across the individual items of the measure, suggesting they cohere sensibly into the overall score.

Ideation

The logic model also identifies the potential for increased creativity in the pupils who have participated in their programme. To explore this, we reported differences in the 'ideation' sub-measure of the WSEM as an additional secondary outcome measure. This measure was jointly chosen with RSA and allows us to address the RQ on pupils' perception of their capacity to generate ideas. This uses the first five questions of the WSEM (see Appendix O for these statements) and possible scores range between 5 and 25 for each child. As this is a sub-domain of the WSEM measure as a whole, data collection details are as per that measure. Also, like the WSEM, the scores are rather negatively skewed (see histogram in Appendix E), with potential consequences for attenuating impact estimates for this outcome.

Sample size

Sample size calculations were carried out taking into account the analysis of the WAM, since this is the primary outcome of interest. The aim was to achieve a minimum detectable effect size (MDES) of *d* = 0.20 with the following assumptions: power of 0.8 for a two-tailed 0.05 significance test, randomisation at school level, an intra-cluster correlation of 0.15[8] (EEF, 2015) and 25 pupils involved in the trial per school with 10 percent pupil-level attrition.

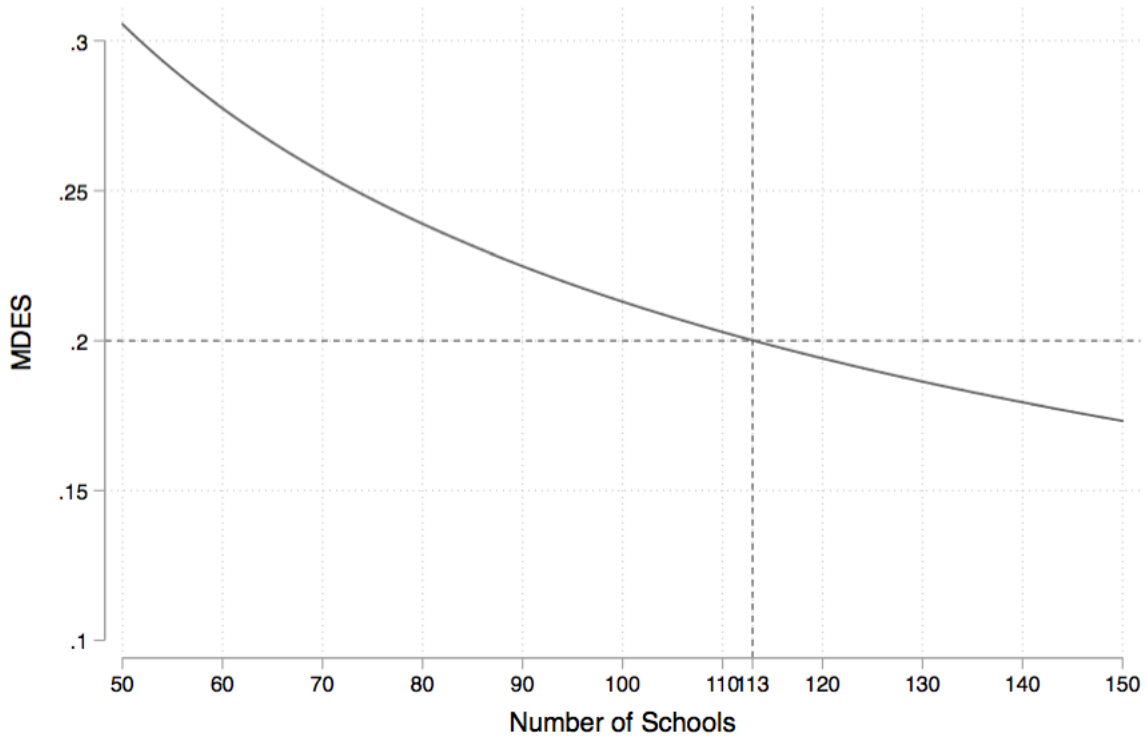
When estimating required effect sizes at the protocol stage, an appropriate pre-test/post-test correlation assumption could not be estimated empirically directly for this trial, since correlation data between the planned baseline and outcome measures were unavailable. This is because the planned baseline measure was the score in the Year 1 phonics screening check (consistent with EEF policy to use an administrative measure rather than an additional pre-test), which has only been in place since 2012, and our primary outcome measure (the WAM) is an even newer measure. EEF guidance suggests that a pre- and post-test correlation of 0.7 in education research is common (Torgerson & Torgerson, 2013); however, we see this as too optimistic in this case. The 21-day test–retest correlation coefficient of the WAM is reported to be 0.82 (Dunsmuir et al., 2015) but the time elapsed between the planned baseline and outcome measure in this trial would have been much longer, and it was never planned to use the WAM itself as a baseline. Our planned baseline measure (score in Year 1 phonics screening check) has less variance than would be ideal, due to a degree of bunching between the pass (32) and highest available mark (40). This bunching presumably occurs in order to push pupils over the pass line due to accountability concerns.

Nevertheless, given the phonics screening check score's closer temporal proximity to the outcome measure point, we expected (and continue to expect) that it is likely to explain more variance in our outcome than earlier measures available in the NPD (which would have to be measured at the EY Foundation Stage). While there is no direct measure of the correlation between the WAM and the phonics screening check available, a value of 0.52 was estimated using Year 1 phonics screening check scores and Progress in International Reading Literacy Study (PIRLS) scores (DfE, 2017) (taken in Year 5, the same year as the WAM will be administered). Given the similar time period between baseline and outcome test administration, and the related domain, we used this estimate as likely to approximate the value that would have been expected in this trial. Based on this, our sample size calculations at the protocol stage assumed that 25% of post-test variance at both pupil- and school-level would be explained by the pre-test (equivalent to pre-test/post-test correlation of 0.5).

The assumptions discussed above suggested a sample of 113 schools in total across both arms (intervention and control) was needed to achieve an MDES of *d* = 0.2 (see Figure 2). Based on discussions between the evaluation team, the PoP team and the EEF at the setup meetings, a sample size of 120 schools was agreed. The PoP team confirmed that recruitment of 120 schools and intervention delivery to 60 treatment schools were reasonable and achievable numbers given their capacity. Conditional on the sample size of 120 schools and the assumptions outlined above, this trial was estimated to be able to detect an effect of 0.18. Assuming the FSM sub-group is 15.3 percent of the total size of the sample (based on pupils aged 5–10 in data from DfE statistics (DfE, 2016) and ignoring that it might have been be higher if recruited schools were in more disadvantaged areas), and maintaining all other assumptions (which is likely to be a conservative approach, given lower levels of within-group variation in this sub-group), the MDES for this group at time of design was estimated to be approximately 0.29 standard deviations.

[8] EEF guidance on ICCs (EEF, 2015) is provided for NPD outcomes. In the absence of ICC data for our outcomes of interest we use this guidance, specifically for the reading fine points score, and, given uncertainty about the geographical spread of participating schools, we use the highest regional ICC (which happens to be Inner London) to the nearest two decimal places.

Figure 2: Minimum detectable effect size (MDES) estimate as a function of number of schools at design stage



Ultimately, 101 schools were recruited to the trial and randomised. As this was below the target level, combined with a small reduction in the estimated average cluster size (from 23 to 22) in the schools recruited, the estimated primary analysis MDES increased to d = 0.21. However, there was a larger number of FSM pupils per school than expected (six pupils per school rather than three, after estimated attrition adjustment) and, as a result, the estimated MDES for this sub-group reduced a little, to d = 0.26.

As noted above, there have been substantial changes to the analysis that we were ultimately able to carry out due to data access issues with our planned baseline measures. These have a substantial bearing on the assumptions that underpin estimations of MDES. Specifically, the change in baseline variables included in the analysis model means that our early assumptions about the proportion of variance explained have ended up being optimistic (although, of course, this could still have been the case even if we had been able to obtain the planned baseline data): our data suggest that with the substituted baseline measures, pupil-level variance explained (R^2) is 0.02 and school-level variance explained is 0.26 (with overall variance explained being 0.06). Furthermore, our average cluster size is lower than anticipated at 20 due to testing challenges, the number of control schools is 43 and the number of intervention schools is 46 (89 schools in total, with a treatment ratio of 0.52), and the intra-cluster correlation is substantially higher than anticipated at 0.33 (compared with 0.15 using standard EEF assumptions). Altogether, this leads to an estimated MDES for the primary analysis of d = 0.34. This implies substantially reduced precision or, equivalently, more uncertainty (larger confidence intervals) around our treatment estimate than would have been the case if our initial assumptions were met. This higher level of uncertainty should be borne in mind when interpreting the findings.

There were similar challenges for the FSM sub-group. There the pupil-level variance explained is 0.01, the school-level variance explained is 0.06 (leading to an overall figure for variance explained of 0.03), the ICC is 0.27, and the average cluster size is 4 (the number of schools are as with the main sample exercise). This results in an estimated MDES for the FSM sub-group analysis of d = 0.40.

Randomisation

Schools recruited by the project team were randomly assigned to treatment and control groups by the evaluation team. Stratification (referred to as blocking in the evaluation protocol) was used to attempt to improve cross-arm comparability of schools and to improve precision of estimates.

Eight strata were defined on the basis of class composition by EAL status (high vs. low split at sample median) and class composition by eligibility for FSM (high vs. low split at sample median). This approach tried to ensure that our treatment and control groups were well balanced in terms of these characteristics, which were likely to be correlated with our outcome measures (EEF, 2015). High and low EAL and FSM schools were defined using the sample median in each case to ensure that block sizes are approximately equal (which may not be the case if we used population, rather than sample, characteristics). Table D7 in Appendix D shows the number of treatment and control schools in each stratum.

Randomisation followed recruitment of schools, including the signing of the MoU and baseline data collection, in March–July 2018. Unlike the other two Learning about Culture KS2 trials, PoP only had one randomisation batch. This randomisation process was conducted using a script run in Stata 14. This followed the following process:

1. The schools were sorted into four strata on the basis of proportion of FSM students (split across the median sample proportion) and proportion of EAL students (split across the median sample proportion).
2. Each school was assigned a randomly generated number (setting a stable seed for the random number generation).
3. The schools were sorted by strata and random number.
4. Schools were assigned to the treatment arm and to the control arm in turn, meaning that even numbered strata would have 50 : 50 allocations to treatment and control but odd numbered strata would have small differences in allocation.

The code that was used to operationalise these steps (including the stable seed to allow for replication of the process) is reported in Appendix F.

Statistical analysis

Primary analysis

Our primary analysis focused on the WAM score and was performed using Stata 15. All continuous variables were used in their 'raw' form (in line with EEF guidance) as there was no clear reason to transform the data.

An ordinary least squares (OLS) model was estimated, in which our outcome variable was regressed on a treatment arm indicator, strata indicators (based on proportion of the class eligible for FSM, proportion of the class identified as EAL, and whether the school was randomised as part of the first or second batch), and, in a deviation from the evaluation protocol, the following baseline variables: indicators of FSM eligibility, EAL status, class-level FSM composition, and class-level EAL composition (further details below).

As noted by EEF guidance, in a model that does not account for clustering, when this is a feature introduced by the experimental design, 'the point estimates will be accurate, but the standard errors will be downward biased' (EEF, 2018, p.3). However, we accounted for the potential effects of the experimental design in this respect by calculating standard errors, taking into account clustering (Angrist & Pischke, 2008) at the school level, which allows for correlation of pupil outcomes within schools. We prefer this to the use of a hierarchical linear model which makes additional assumptions about the school-level effects that may not be justified. We also estimated randomisation–inference *p*-values accounting for the clustering and stratification of the design, which were consistent with those based on clustered sampling inference (see Table D4 in Appendix D).

The estimated impacts are intention-to-treat (ITT) effects and have been reported with 95% confidence intervals. Intra-cluster correlations have also been reported.

In the evaluation protocol and statistical analysis plan (SAP), we stated our intent to estimate the following model in order to estimate the ITT impact of the intervention:

$$Y_{ij} = \alpha + \beta_1 Treat_j + \beta_2 PreTest_{ij} + \gamma' X_j + \varepsilon_{ij},$$

where individual i is nested in school j, $Y_{ij}$is the WAM score, $PreTest_{ij}$ is the value of the phonics screening check score (using the NPD variable PHONICS_MARK) used as a pre-test, $Treat_j$ is our school-level treatment indicator, X_j is a vector of stratification variables, and ε_{ij} is an error term. Standard errors are calculated allowing for clustering at school level (j).

However, because of the data access issues discussed above, we were unable to estimate this model due to the unavailability of PHONICS_MARK as $PreTest_{ij}$. Instead, we estimate the following model in which $PreTest_{ij}$ has been replaced with FSM eligibility, EAL status, class-level FSM composition and class-level EAL composition (as discussed above):

$$Y_{ij} = \alpha + \beta_1 Treat_j + \beta_2 FSM_{ij} + \beta_3 EAL_{ij} + \beta_4 FSMProp_j + \beta_5 EALProp_j + \gamma' X_j + \varepsilon_{ij},$$

where everything is as per the planned model except that FSM_{ij} is whether individual i is eligible for FSM and, similarly, EAL_{ij} is whether individual i is recorded as having EAL, while $FSMProp_j$ is the FSM composition of treated class in school j, and $EALProp_j$ is the same for its EAL composition.

As such, this report's primary ITT estimate is recovered from the estimate of β_1 in this latter model when it is estimated on the full sample at randomisation.

Note that while this model is a deviation from the evaluation protocol and SAP, it was planned and reported to the EEF ahead of analysis being carried out. The model has not been altered depending on the significance of any variables included (i.e., no variables were removed due to being statistically insignificant) including the vector of blocking variables (X_j). (Syntax for this primary analysis model is reported in Appendix F.)

Secondary analysis

We conducted two secondary outcome analyses.

- **Writing self-efficacy**: Same as the revised primary outcome analysis, except replacing Y_{ij} with the WSEM score. Note that this was a deviation from protocol, which stated that this would be the same as the planned primary outcome analysis except replacing Y_{ij} with the WSEM score and $PreTest_{ij}$ with assessment of pupils' personal, social and emotional development skills from the Early Years Foundation Stage Profile (aggregated scores from NPD FSP_PSE_G06, FSP_PSE_G07 and FSP_PSE_G08). This change was made due to data access problems rendering these NPD variables unavailable.
- **Ideation**: Same as the revised primary outcome analysis except replacing Y_{ij} with the Ideation sub-score from the WSEM . Note that this was a deviation from protocol, which stated that this would be the same as the planned primary outcome analysis, except replacing Y_{ij} with the Ideation sub-score from the WSEM and $PreTest_{ij}$ with assessment of pupils' personal, social and emotional development skills from the Early Years Foundation Stage Profile (aggregated scores from NPD FSP_PSE_G06, FSP_PSE_G07 and FSP_PSE_G08). This change was made due to data access problems rendering these NPD variables unavailable.

It was also planned, potentially as part of a separate report, to estimate the impact on KS2 grammar, punctuation and spelling test attainment. Unfortunately, the relevant KS2 national curriculum tests in summer 2020 that would have collected these data were cancelled as a result of the COVID-19 pandemic. As such, this medium-term follow up is no longer possible.

Analysis in the presence of non-compliance

The following criteria have been defined in the trial protocol as variables that can be used to assess compliance with the intervention. This draws principally on attendance data collected from the project team. Compliance was measured at the teacher level, which reflects the intervention delivery method. A school will be considered to have complied if and only if the following three conditions are met:

- Teacher attends all 2.5 days training sessions;
- Teacher delivers at least 50% of lessons in both block one and two (10/20);
- Teacher's class attends the half day author off-site workshop.

We used complier average causal effect (CACE) analysis (Gerber & Green, 2012) to estimate intervention effects on treated children. We estimated the CACE using two-stage least squares (2SLS) regression by estimating a (first stage) model of compliance, as follows:

$$Comply_j = \alpha + \beta_1 Treat_j + \gamma' X_j + \xi_{ij},$$

where $Comply_j$ is the binary compliance variable defined above, and ξ_{ij} is an error term. The predicted values of $Comply_j$ from the first stage are used in the estimation of a (structural) model of our outcome measure Y_{ij}. In other respects, the specification remains the same as the revised primary outcome ITT model. This second stage model is specified as follows:

$$Y_{ij} = \alpha + \beta_1 \widehat{Comply}_j + \beta_2 PreTest_{ij} + \gamma' X_j + \omega_{ij},$$

where $\widehat{Comply}_j$ are the predicted values of treatment receipt derived from the first stage model, and ω_{ij} is an error term. Our primary outcome of interest is β_1, which should recover the effect of the intervention among compliers.

We conducted this analysis using the 'ivregress' functionality of Stata to make necessary adjustments to standard errors (which have also been clustered at school level) due to the instrumental variables approach. We note the deviation to protocol due to these models being based on the revised primary outcome ITT model, rather than the planned primary outcome ITT model, which is for the same underlying reasons of data access.

Missing data analysis

In this section we describe and summarise the extent of missing data in the primary and secondary outcomes, and in the model associated with the analysis. Reasons for missing data are also described.

For all models we planned to implement a missing data strategy if more than 5% of data in the model was missing or if more than 10% of data for a single school was missing. The strategy would have been followed separately for each instance of model and variable for which the threshold was exceeded.

In implementing the planned strategy, we first assessed whether the missing data was missing at random (MAR), since this is a prerequisite for missing data modelling to produce meaningful results. To do this we created an indicator variable for each variable in the impact model, specifying whether the data was missing or not. We then used logistic regression to test whether this missing status could be predicted from the following variables: all variables in the analysis model plus eligibility for FSM (and proportion eligible for FSM in the school), and EAL status (and proportion EAL in the school). Where predictability was confirmed we proceeded to the appropriate next step of this strategy.

For situations for which the MAR assumption appeared to hold and only the outcome variable in the model was missing, we re-estimated the treatment effect using our pre-specified model, with the addition of the covariates found to be statistically significantly predictive of missingness of the outcome.

For situations for which the MAR assumption appeared to hold and any variable other than the outcome variable in the model was missing, we used all variables in the analysis model plus eligibility for FSM (and proportion eligible for FSM in the school), and EAL status (and proportion EAL in the school) to estimate a multiple imputation (MI) model using a fully conditional specification, implemented using Stata MI to create 20 imputed data sets. We re-estimated the treatment effect using each data set, took the average and estimated standard error using Rubin's combination rules (Rubin, 2004).

Analysis that is altered following the missing data strategy (either on a multiply imputed data set or with additional variables) would only ever be viewed as a sensitivity analysis. As such, the main estimates of the effectiveness of the treatment are derived from complete case analysis only. However, the sensitivity of the estimates to missingness would be assessed by comparing this main analysis with those altered following the missing data strategy. For example, if the complete case analysis model were to imply effectiveness but the imputed estimate did not, we would assume that the missing data is missing systematically to such an extent as to invalidate the initial conclusion of effectiveness, which would then be stated in the evaluation.

Sub-group analyses

Following EEF guidance, we first tested for an interaction of the treatment and FSM status. This was originally planned to be carried out using the NPD variable EVERFSM_6_P (in line with EEF guidance) and the following model:

$$Y_{ij} = \alpha + \beta_1 Treat_j + \beta_2 FSMEver_{ij} + \beta_3 Treat_j * FSMEver_{ij} + \beta_4 PreTest_{ij} + \gamma' X_j + \nu_{ij},$$

where individual i is nested in school j, $Y_{ij}$is the WAM score, $PreTest_{ij}$ is the value of the phonics screening check score used as a pre-test, $Treat_j$ is our school-level treatment indicator, $FSMEver_{ij}$ is an indicator of FSM eligibility

(EVERFSM_6_P), $Treat_j * FSMEver_{ij}$ is an interaction between these two terms, X_j is a vector of stratification variables and v_{ij} is an error term.

However, in a deviation from protocol, due to the unavailability of the NPD derived indicator of FSM eligibility (EVERFSM_6_P) and the unavailability of the phonics screening check score (for data access reasons, as discussed above), this model was adapted as follows:

$$Y_{ij} = \alpha + \beta_1 Treat_j + \beta_2 FSM_{ij} + \beta_3 Treat_j * FSM_{ij} + \beta_4 EAL_{ij} + FSMProp_j + EALProp_j + \gamma' X_j + v_{ij},$$

where all terms are defined as per the planned FSM interaction model above or the planned primary analysis model.

Standard errors have been calculated allowing for clustering at school-level (j). Syntax for this interaction model is reported in Appendix F.

If a significant interaction was found (i.e., the absolute value of the point estimate of β_3 divided by the school-level clustered standard error is greater than 1.96), then we would have proceeded to conduct a specific sub-group analysis for those who are identified as eligible for FSM ahead of randomisation. This would have been done using the same model as our revised primary analysis. We note that this is also a deviation from protocol, as we would have defined this sub-group as those who have ever been registered for FSM in the NPD (identified using the variable EVERFSM_6_P) and used our planned primary analysis model.

But for the small difference in definition of FSM (which investigation with the DfE suggests is unlikely to be material as no cleaning of the data submitted by schools is carried out before it is made available in the NPD), this sub-group was identified in the trial protocol and FSM pupils are a key sub-group to be analysed in all EEF trials.

This FSM sub-group analysis was conducted for both the primary and secondary outcomes.

Additional analyses and robustness checks

No additional analyses were planned as part of the project's SAP. All additional analyses and robustness checks carried out should be considered exploratory only.

In addition to carrying out inference through school-level clustered standard errors, we also estimated randomisation inference *p*-values in order to check the robustness of inference to this approach. As this was not planned in the SAP, it should be considered exploratory and so it will not be used to guide interpretation of the results. However, it will provide useful information on the extent to which there is variation between these different approaches to statistical inference.

We run three exploratory robustness check models based on potential issues identified in the course of analysis:

- Due to delays in the testing of some schools, we ran a robustness check model in which we replicated the primary analysis model but added a control for the number days between the date that the first school was tested and the date that the school in question was tested. The logic for this is that delays to testing could have affected the dosage of the intervention.
- It is possible that variation in the approach of different markers (e.g., a degree of leniency by some markers, despite the steps documented above to maximise consistency) who marked the WAM (primary outcome measure) could affect the treatment estimate. While marking was blind to treatment assignment, the relatively small number of markers could lead to imbalance in such approaches by chance. As such, we ran a robustness check model in which we replicated the primary analysis model but added marker fixed effects.
- Imbalance in school level KS1 scores between treatment and control group schools in the analysis sample is identified as part of our balance checks. To check whether such imbalance might explain our findings, we ran a robustness check model which replicates the primary analysis model but includes average school level KS1 scores as an additional covariate. As it was not possible to link in average KS1 scores for all schools, multiple imputation was used for this analysis, carrying out 20 regression-based imputations of average KS1 score using all covariates in the primary analysis model.

Estimation of effect sizes

Hedges' *g* effect size was calculated as set out by Hedges (1981):

$$g = J(n_1 + n_2 + 2)\frac{x_1 - x_2}{\widehat{s^*}},$$

where our conditional estimate of $\underline{x_1} - \underline{x_2}$ is recovered from $\beta_1$in the primary ITT analysis model;

$\widehat{s*}$ is estimated from the analysis sample as follows:

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

where n_1 is the sample size in the control group, n_2 is the sample size in the treatment group, s_1 is the standard deviation of the control group, and s_2 is the standard deviation of the treatment group (all estimates of standard deviation used are unconditional, in line with the EEF's analysis guidance to maximise comparability with other trials);

and $J(n_1 + n_2 + 2)$ is calculated as follows:

$$J(n_1 + n_2 + 2) = \frac{\Gamma\left(\frac{n_1+n_2+2}{2}\right)}{\sqrt{\frac{n_1+n_2+2}{2}}\,\Gamma\left(\frac{n_1+n_2+2-1}{2}\right)},$$

where n_1 is the sample size in the control group and n_2 is the sample size in the treatment group or, if calculating $J(n_1 + n_2 + 2)$ proves computationally intractable[9] using the above method, we would instead fall back on the following approximation:

$$J(n_1 + n_2 + 2) \approx (1 - \frac{3}{4(n_1+n_2)-9}).$$

Ninety-five percent confidence intervals (95% CI) of the effect size have been estimated by inputting the upper and lower confidence limits of $\widehat{\beta_1}$ from the regression model into the effect size formula.

Estimation of intra-cluster correlation (ICC)

In order to estimate the ICC of the outcome measure at school level we employed an empty variance components model, as follows:

$$Y_{ij} = \alpha + \eta_j + \varepsilon_{ij,}$$

where individual i is nested in school j, $Y_{ij}$is the WAM score, η_j is a school-level random effect, and ν_{ij} is an individual-level error term. The school-level random effect is assumed to be normally distributed and uncorrelated with the individual-level errors. An empty variance components model is used to facilitate comparability between trials (and in line with EEF guidance).

The ICC was then estimated from this model using the equation:

$$\rho = \frac{var(\eta_j)}{var(\eta_j) + var(\varepsilon_{ij})}.$$

In the SAP, we also intended to estimate the ICC of the planned baseline measure, the phonics screening check score. In a deviation from this plan, this analysis was not conducted for reasons of data access, as discussed above.

Longitudinal analysis

We had planned to estimate the effect of the intervention on pupil performance in KS2 national curriculum tests in English grammar, punctuation and spelling. Unfortunately, the relevant KS2 national curriculum tests in summer 2020 that would have collected these data were cancelled as a result of the COVID-19 pandemic. As such, this medium-term follow up is no longer possible. It may ultimately be possible to consider analysing impacts on GCSE results (e.g., GCSE English, taken in 2025); however, planning for this is beyond the scope of this report.

[9] The output of the gamma (γ) function in the Hedges' *g* correction factor (*J*) becomes large quickly, making this method of computation intractable where $n_1 + n_2$ is not small. As such, it can quickly become intractable. Thankfully, the approximate method tends towards the full correction factor quickly. As such, where the computational intractability is an issue, the approximate method is appropriate. In any event, the correction factor is likely to be small in this trial.

Implementation and process evaluation (IPE)

As part of the mixed method design of this evaluation, an IPE was conducted to complement the findings from the impact evaluation. The IPE involved members of the team with expertise and knowledge of the arts and education, which they fed into the design, conduct and analysis of the IPE. This section describes the IPE aims, data collection and analysis methods used.

IPE research questions

The purpose of the IPE was to understand how teachers experienced the PoP training, how teachers delivered PoP lessons in their classroom and, in particular, to better understand the barriers and facilitators to implementation and delivering the intervention with fidelity. The IPE was also used to identify how teaching through PoP compares with usual practice, and to determine the cost to schools of delivering the intervention. (See the *Evaluation objectives* subsection for the full list of the IPE research questions.)

Multiple sources of data were triangulated to address the IPE research questions. The primary approach to IPE data collection consisted of collecting case study data from six schools in the intervention arm of the study. Each case study involved the following methods: a semi-structured interview with the teacher involved in delivering the intervention; a semi-structured interview with a member of the SLT; informal discussions with pupils; and an observation of a PoP-style lesson delivered by the interviewed teacher. A survey was also sent to all intervention and control group schools and administrative data on lesson delivery were collected by teachers in intervention schools. In addition, two author-illustrator workshops (November 2018) and two continuing professional development (CPD) sessions (October 2018 and January 2019) were observed. The PoP team also collected data on attendance at training sessions and the author-illustrator workshops. Intervention manuals and guidance were also reviewed to inform interviews and data analysis.

The research questions and the data collection methods used to answer them are shown in Table 3.

Table 3: IPE methods overview

Data collection method (case studies)	Data analysis method	Participant groups	Target number of participants per case study	Actual number of participants/activities per case study	Total number of participants/ activities	Research questions addressed
Semi-structured Interviews	Framework approach	SLT	1	1	6	RQ1 (Fidelity), RQ3 (Quality), RQ5 (Home), RQ6 (Teacher change), RQ9 (Mechanisms)
	Framework approach	Teachers	1	1	6	RQ1 (Fidelity), RQ2 (Responsiveness), RQ3 (Quality), RQ5 (Home), RQ6 (Teacher change), RQ8 (Engagement), RQ9 (Mechanisms)
Informal group interview	Framework approach	Pupils	2	2–9 per school	At least 27[1]	RQ8 (Engagement)
Observations	Framework approach	Pupils and teachers	1 lesson	1 lesson (except one school)	5 lessons	RQ1 (Fidelity), RQ8 (Engagement)
Observations	Framework approach	Author-illustrator, teachers	1 author-illustrator training teachers; 1 mid-point training	2 observations as described, additionally 2 observations of author-illustrator teaching pupils		RQ1b (Training)
Baseline surveys	Descriptive statistics	Treatment teachers	51	21		RQ1 (Fidelity), RQ1a (SLT), RQ6 (Teacher change)
Follow-up surveys	Descriptive statistics	Teachers	51 treatment, 50 control	43 treatment, 41 control		RQ1c (Effective), RQ5 (Home), RQ8 (Engagement), RQ6 (Teacher change)
Cost interview	Framework approach	Teachers	4	4		Cost

[1] Exact number of pupils interviewed not recorded at two schools, but at least two were interviewed.

Sampling and recruitment

Case studies

Six case study schools were selected using a purposive sampling approach, to capture the range of Year 5 classes that had received the PoP intervention during the 2018/2019 academic year. The primary sampling criteria were: (1) school engagement in the intervention (defined as high where schools had delivered all 20 PoP lessons, and low where schools had only delivered 10 lessons at the point of sampling); and (2) the proportion of pupils receiving FSM (defined as high or low depending on whether the school was higher or lower than the median for all PoP intervention schools, which was 12.4% [information obtained from UK Government; 2019]). The secondary sampling criteria were: (1) the author-illustrator involved in delivery at the school; (2) geographical region (categorised as East of England, Greater London, North East, North West or South East of England); and (3) Ofsted rating (recorded as Outstanding, Good or Requires improvement [information obtained from UK Government, 2019]). Table 4 sets out the characteristics of the six case study schools that were recruited.

Table 4: Characteristics of case study schools

Characteristic	Number
Number of Power of Pictures (PoP) lessons delivered	
20	3
10	3
Free school meals (FSM)	
0–5%	0
6–10%	2
11–15%	1
16–20%	1
21–25%	1
26–30%	1
Author-illustrator involved in delivery	
Alexis Deacon	2
Tim Hopgood	2
Viviane Schwarz	2
Geographical region	
Greater London	3

North East	2
South East	1
Ofsted rating	
Outstanding	2
Good	2
Requires Improvement	2

Surveys

Baseline and follow-up surveys were administered to schools taking part in the trial (see Appendix K and Appendix L for an overview of survey responses). The baseline survey was distributed to schools randomised to the intervention arm of the trial in October 2018. Unfortunately, the response rate (calculated using the number of schools originally randomised to the intervention arm of the trial) was very low (41%, 21 of 51 schools), so it did not produce meaningful data. Instead, the baseline survey questions were incorporated into the follow-up survey. For the follow-up survey, a census approach was taken, meaning that all schools involved in the trial were invited to respond. Depending on the arm of the trial to which they were randomised, schools were sent either an intervention or a control survey. Because we are not conducting any statistical inference with the survey results, confidence intervals are not given for response frequencies; in addition, as the schools taking part in the trial were not a representative sample of all primary schools in the UK, it is not appropriate to generalise the findings beyond this group of schools. Because we are not conducting any statistical inference with the survey results, confidence intervals are not given for response frequencies; in addition, as the schools taking part in the trial were not a representative sample of all primary schools in the UK, it is not appropriate to generalise the findings beyond this group of schools.

Data collection

Case studies

Sampled schools were contacted by email and, where schools agreed to take part, a date was arranged for a researcher to visit. All visits took place between May and June 2019. At the visit, data was collected via (1) a semi-structured, audio-recorded interview with the Year 5 teacher involved in the intervention; and (2) a member of the school's SLT; (3) notes were taken during discussions with pupils; and (4) the interviewed teacher was observed delivering a PoP-style lesson. The exception to this was one school where they had finished all PoP-related teaching so an observation could not be carried out.

In addition, it was originally envisaged in the protocol that interviews would be conducted with teachers both before and after the observation. The first interview was intended to focus on their broader experience of the programme; the second interview was intended to discuss the observation. However, when arranging interviews, teachers expressed concern at leaving their classroom twice or had a timetable that did not support conducting interviews before and after instruction; instead, it was agreed with teachers that they would attend one summative interview. Given the design of the interview schedule, this change did not pose any risk to the quality of the data being collected.

The interviews were conducted using a topic guide that explored what usual practice looks like in relation to teaching with picturebooks, the context in which the intervention was implemented, how the training was translated into classroom teaching, the facilitators and barriers to implementation, children's engagement, the perceived impact of the intervention and the mechanisms underlying this change. Discussions with children covered what they had done in PoP lessons and how they had found them, their experience of attending the author-illustrator workshop, and how PoP compares with other teaching of writing and drawing. Interviewees were informed that the interview was anonymous, that they could withdraw at any time, and that they did not have to answer any questions they did not want to. (Full topic guides can be

found in Appendices G–I.) Observations were used to capture the teacher's approach to delivery, the content of the lesson and children's engagement with the lesson. The observations proforma was developed collaboratively in partnership with the UCL team, who brought their respective subject matter expertise. Specifically, Dominic Wyse has writing, music and mixed-methods research expertise; Gemma Moss has expertise in early literacy development; and Andrew Burn has expertise in English, media and drama. Observation notes were recorded on a structured proforma (see Appendix J), and primarily used to help researchers probe effectively during the interviews and to deepen understanding of interview findings. Strategies to reduce bias were implemented, including having two researchers (e.g., a BIT and UCL team member) conducting the initial observations, discussing and agreeing upon the final observation notes that were recorded; one of those researchers continued to conduct the subsequent observations in the other case study schools. The researchers did not inform the delivery team about which schools were visited.

Surveys

The baseline survey was informed by the programme logic model and was designed to capture treatment school buy-in and teacher attitude towards the programme. Data from interviews, as well as feedback from the delivery team, was used to inform the design of the follow-up surveys. The survey administered to intervention schools covered experiences of training and barriers to attendance, confidence in delivering PoP lessons, views on the author-illustrator involved in delivering PoP, and the perceived impact and quality of the intervention. The control survey focused on understanding usual practice in relation to teaching using picturebooks. Full details on the intervention and control group surveys can be found in Appendices K and L.

A link to the online survey platform SmartSurvey was sent to the key contacts at each school in June 2019. For the intervention survey, the email stated that the survey should be completed by the teacher who had been involved in the PoP intervention, while the control survey email stated that a Year 5 teacher should complete the survey. Reminders were sent to schools who did not initially complete the survey. When schools did not respond to reminders, these were followed up with a phone call and teachers were given the opportunity to complete the survey over the phone. All data was collected by the end of July 2019.

Administrative data

Data was collected on the number of training sessions attended by the class teacher (out of a total of 2.5 days) and the number of lessons delivered by teachers in block one and two (out of a total of 20). This data was sent once by the delivery partner in April 2019 to inform case study sampling, and again at the end of the trial when the complete data set was available, as a measure of school engagement with the intervention. As the number of training sessions attended varied minimally between schools, only the number of PoP lessons delivered was used to inform case study sampling.

IPE analysis

Case studies

Verbatim transcripts of the interviews and notes from the observations were analysed using the Framework approach (Ritchie et al., 2013). Firstly, emerging themes were identified through familiarisation of the data. The analytical framework was then created using a series of matrices in Excel, each relating to an emergent theme. The columns in each matrix represented the key sub-themes drawn from the findings, and the rows represented individual participants interviewed or schools observed.

The interview and observation data were then summarised in the appropriate cell, which meant that all data relevant to a particular theme was noted, ordered and accessible, facilitating a systematic approach to analysis that was grounded in participants' and schools' accounts. Analysis involved working through the charted data to draw out the range of schools' experiences and participants' views and identifying similarities, differences and links between them. Thematic analysis (undertaken by looking down the theme-based columns in Framework) identified concepts and themes, and the case-based analysis (undertaken by comparing and contrasting rows in Framework), enabled links within cases to be established and cases compared and contrasted with each other.

During the analytical process a balance was maintained between deduction (using existing knowledge and the research questions to guide the analysis) and induction (allowing concepts and ways of interpreting experience to emerge from the data). The IPE findings section is organised based on the identified themes and sub-themes, which are outlined and described in their respective subsections in the IPE findings. Multiple strategies were employed by the researchers to increase the credibility (i.e., accurate representation of the data), transferability (i.e., potential to apply the findings to

other settings), dependability (i.e., traceable, logical analytical process) and confirmability (i.e., being grounded and traceable to the raw data) of the findings, with the ultimate aim of reducing the bias during the analytical process (Hannes, 2011).

First, in terms of striving to increase credibility, the researchers conducted peer debriefing meetings with the senior qualitative research lead (Matt Barnard, Head of Evaluation at BIT) and qualitative researchers who were not directly involved in the data collection or analysis process for the respective intervention. In addition, in accordance with the chosen approach to data analysis, the researchers focused on describing range and diversity, including the noting of any disconfirming cases. Verbatim participant quotations are used to provide evidence and exemplify the theme(s) discussed in the paragraph before the quotation. Quotations were selected by the qualitative researchers who conducted the data analysis, by considering multiple factors, including how well they exemplify the theme(s) discussed. The researchers also sought to ensure that the quotations used in the IPE findings capture the variation in terms of points of view and experiences, as well as types of participant interviewed (e.g., SLT, teachers) and the associated schools. Further, as qualitative data can only be generalised in terms of range and diversity and not in terms of prevalence, the analytical outputs focus on the nature of experiences, avoiding numerical summaries or language such as 'most' and 'majority'.

Second, to increase the potential for transferability and assessment of applicability to other contexts, the *Context and motivations for participation* subsection of the *Implementation and process evaluation (IPE) results* section describes the key details of the case study schools and the selection criteria. The *Implementation and process evaluation (IPE) results* section also includes descriptions of the co-facilitators (e.g., years of teaching and drama experience) and important contextual details about the case study schools (e.g., motivations for choosing to implement the intervention).

Third, to increase dependability and confirmability of the findings, the researchers maintained a detailed audit trail and triangulated the data by comparing the findings from multiple types of participants (e.g., teachers, delivery partners) and sources of data (e.g., survey, interviews, observations). Researchers adhered to the key principles of the Framework approach, which includes ensuring that data management and analysis is systematic, comprehensive, transparent and grounded in the participants' accounts. Doing this was facilitated by the creation of a series of matrices in Excel that contained descriptive summaries of data that can be easily traced back to the verbatim quote on the relevant page of the transcript being described.

Surveys

Follow-up survey data was first cleaned by identifying schools that had returned two or more survey responses. In these instances, the most recent survey completed by a respondent who identified as a 'teacher', was used for analysis. For instance, if there were two responses, both from teachers, then the most recent one was kept, whereas if one response was from a member of SLT and one was from a teacher, the SLT response was removed even if this was the more recent. Responses from teachers were prioritised where there were multiple responses from the same school because it was assumed that the class teacher would have been more involved in the delivery of the intervention (for the intervention survey) and know more about usual classroom practice (for the control survey), and therefore, their views and experiences were most relevant. Data was also removed from one intervention school respondent who repeatedly put 'don't know' and gave answers indicating that they had not been involved with PoP and would therefore have limited information to provide about the intervention. Due to low response rate, baseline survey data were not analysed.

Prior to cleaning, there were 44 intervention school responses and 49 control school responses. Following cleaning, there were 43 responses from intervention schools (out of 51 schools randomised to the intervention group) and 41 responses from control schools (out of 50 schools randomised to the control group), giving a response rate for intervention schools of 84% and for control schools of 82%. Stata (version 14) was used to generate descriptive statistics for each question. Percentages scores are reported, where relevant, in the IPE findings section. Complete survey findings are provided in Appendices K and L.

Costs

The evaluation gathered three key categories of data: direct marginal costs (which will form the basis of the cost per pupil); pre-requisites (which will be reported separately from the cost per pupil); and school staff time. The data was gathered in two ways. Firstly, the evaluators requested information from delivery partners on how much they charged

schools for delivering the intervention as part of the evaluation and how much they will charge schools in the future, excluding any funding or subsidy that is associated with delivering the intervention as part of this evaluation. The latter data is used in calculating the cost per pupil; the former data is to ensure there is clarity about the precise nature of the data that is being requested and transparency of the approach.

The second mode of data collection was the use of case study interviews, as specified in the protocol. The interviews were used to determine whether questions about costs would be included in surveys, with the decision also taking into account survey length and risk of damaging response rates. Based on low responses to the baseline survey, it was judged appropriate to omit cost-related questions so as to keep survey length down and not potentially dampen response rates. Instead, costs were further explored through interviews, which were also deemed a more appropriate method to gather detailed data, as they allow for follow-up questions to clarify responses and probe for more information.

The evaluation team felt the programme cost was best estimated by having a good sense of the range and diversity of experiences, which was facilitated by using a case study approach supported by purposive sampling (Ritchie et al., 2013). Case study schools were selected from those who had good engagement with the programme, as they were more likely to give the best indication of the resources needed to implement the programme fully; including schools with little engagement was likely to artificially deflate costs. As an indicator of this, we selected schools from the pool who had completed outcome data collection (as fidelity data was not available at that stage).

The resources required to deliver the intervention were most influenced by staff time and any related marginal costs (such as travel and subsistence). The evaluators assumed these things were most likely to be related to the amount a school has to spend per pupil and the nature of the local area and school population, the most relevant indicator for which is the percentage of pupils eligible for FSM. That is, school spending on the programme is likely related to school financial resources. Therefore, one school was purposively selected in each of the following four categories:

1. School with percentage of FSM in top half of participating schools and spend per pupil in top half;
2. School with percentage of FSM in top half of participating schools and spend per pupil in bottom half;
3. School with percentage of FSM in bottom half of participating schools and spend per pupil in top half;
4. School with percentage of FSM in bottom half of participating schools and spend per pupil in bottom half.

The case study data was collected by RAs employed and trained by BIT. The RAs conducted interviews with teachers via telephone, using a structured interview guide designed by BIT for this purpose. RAs estimated it took approximately 20 minutes to complete the discussion.

Teachers were asked to report on direct costs of the intervention to the school, materials purchased, travel and subsistence, the cost of covering staff at training and the cost of any new physical materials purchased to improve the classroom environment. Teachers also reported on time spent embedding the intervention in their school, time at training, as well as time spent preparing to deliver the intervention. Staff were also asked to report on time taken to organise supply cover and the amount of supply cover.

Data from these interviews were used to calculate the financial and time costs outlined in this report.

Timeline

Table 5: Timeline

(a) Overall project timeline

Dates	Activity	Staff responsible / leading
June 2015	Participants sit Year 1 Phonics screening check, which will be used as baseline measure.	N/A
November 2017–February 2018	Recruitment: The PoP team began recruitment halfway through the autumn term 2017 using its existing network of schools.	PoP team

	This had to be extended substantially into 2018 due to challenges with recruitment.	
October 2017–September 2019	Pre-randomisation data collection	BIT and PoP team
July 2018	Randomisation	UCL
October 2018–June 2019	Intervention in schools	PoP team
October 2018–June 2019	Implementation and process evaluation (IPE) fieldwork (see further detail on IPE timeline below)	BIT and UCL
May–July 2019	Outcome testing: Pupils' writing and self-efficacy outcomes will be collected by BIT. These tasks will be marked by PGCE students at UCL in a process overseen by BIT.	BIT and UCL
July–September 2019	Collation and cleaning of outcomes and compliance data in readiness for upload to ONS SRS for linkage with DfE NPD extract	UCL and BIT
January–July 2020	Project paused awaiting conclusion of data sharing agreement necessary for upload of project data to the ONS SRS for linkage with NPD.	
August–October 2020	Project resumed with planned revisions as a result of delays in achieving data sharing agreement. Impact analysis and report writing. UCL led on the data analysis with agreed deviations from published statistical analysis plan (SAP).	UCL and BIT

(b) IPE timeline

Date	Item
Autumn term 2018	Observation of Illustrator training
	Collection of baseline survey to measure school buy-in and teacher attitude towards intervention
	Collection of school characteristics
Spring term 2019	Observation of mid-point training
	Collection of fidelity data to inform case study sampling
	Finalised sampling strategy
	Conducted in-school case studies
Summer term 2019	Conducted in-school case studies
	Administered end of intervention survey
	Conducted analysis

Impact evaluation results

Summary

- There was a small and statistically insignificant impact of participating in the PoP on writing attainment (WAM). This effect size of 0.09 is equivalent to one month's progress, but due to uncertainty could be between minus one month's progress and plus one month's progress.
- There was suggestive evidence of a positive and marginally significant impact of participating in the PoP on writing self-efficacy (WSEM), but no effect on ideation. The effect size on the WSEM, 0.11, is equivalent to two months' progress, but due to uncertainty is only statistically significant at the 10 percent significance level.
- There was no differential effect of the intervention for FSM-eligible pupils. Here, the effect size was 0.06, which is equivalent to one month's progress, but due to uncertainty could be between minus one month's progress and plus one month's progress.

Participant flow including losses and exclusions

The flow of participants is detailed in Figure 3. Of the original 484 schools that were approached, 302 did not respond, three did not meet the inclusion criteria because they were independent schools, and 78 declined to participate for other reasons (notably concerns around GDPR and finding it challenging to submit pupil data in the required timeframe). In total, 101 schools agreed to participate in the trial, met the eligibility criteria, and provided the necessary baseline pupil data for randomisation to proceed. These remaining 101 schools were randomly allocated to the intervention and control groups using a block stratified randomisation, as described above. At randomisation, 1264 students in 51 schools were allocated to the intervention group and 1410 students in 50 schools to the control group.

Five treatment and seven control schools declined to participate in testing at the end of the intervention, and some pupils were absent at the point of testing (and mop-up testing was not successful), such that outcome data was not collected for 262 students in the treatment group and 467 students in the control group. This meant that, ultimately, 1002 pupils in 46 schools allocated to treatment and 943 pupils in 43 schools allocated to control were analysed.

The MDES estimated at various points of the trial is reported in Table 6. This was 0.18 at the design stage, increased to 0.21 at randomisation (primarily due to under-recruitment), and increased again to 0.32 at the stage of analysis. This increase may be attributed to a combination of being unable to access the planned baseline measures (discussed further above), attrition rates and a higher intra-cluster correlation of the outcome measure than was anticipated based on previous EEF guidance.

Figure 3: Participant flow diagram

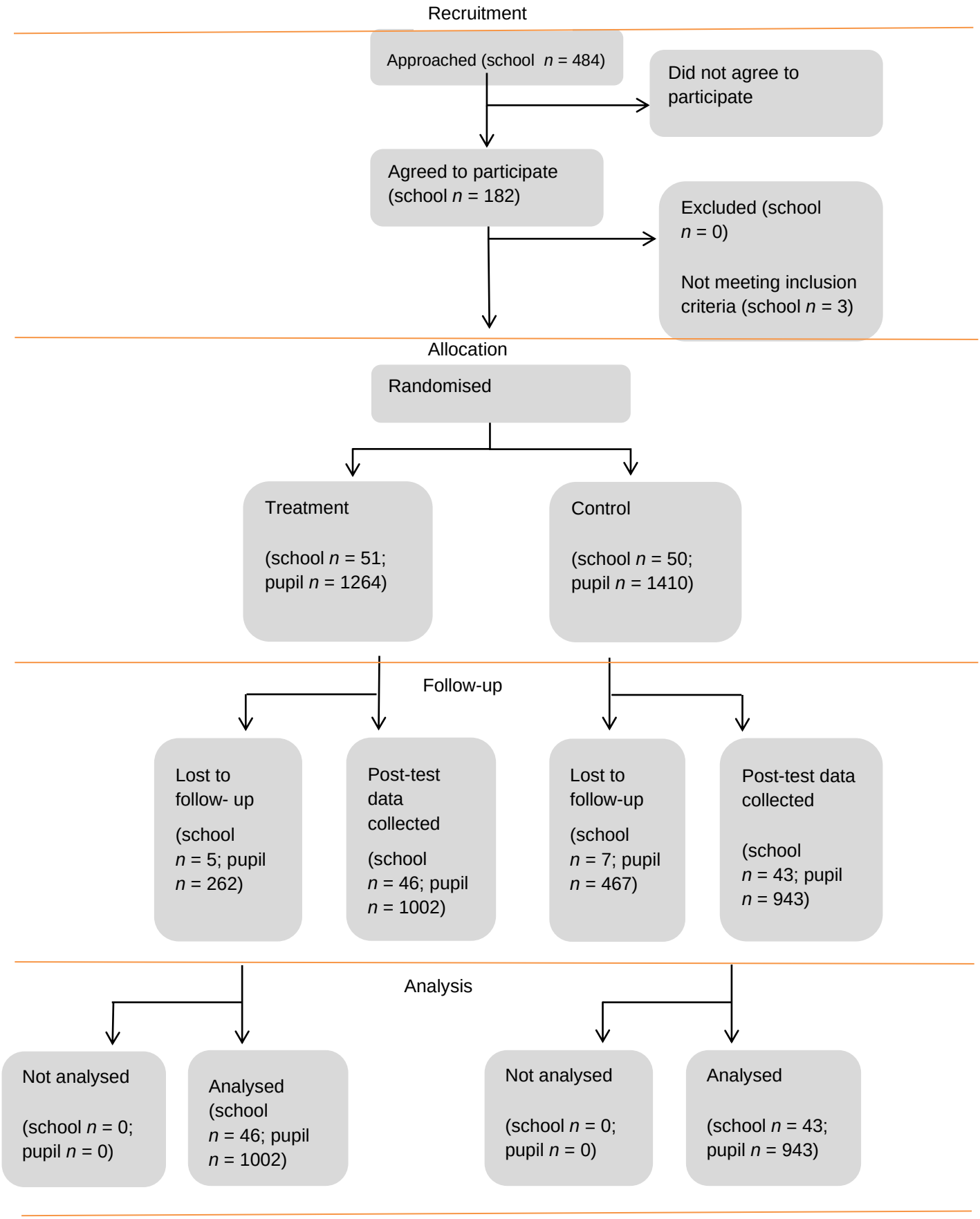


Table 6: Minimum detectable effect size (MDES) at different stages

		Protocol		Randomisation		Analysis	
		Overall	FSM	Overall	FSM	Overall	FSM
MDES		0.18	0.29	0.21	0.26	0.32	0.40
Pre-test/post-test correlations	Level 1 (pupil)	0.50	0.50	0.50	0.50	0.14	0.10
	Level 2 (class)	0.50	0.50	0.50	0.50	0.47	0.24
	Level 3 (school)	N/A	N/A	N/A	N/A	N/A	N/A
Intra-cluster correlations (ICCs)	Level 2 (class)	0.15	0.15	0.15	0.15	0.31	0.27
	Level 3 (school)	N/A	N/A	N/A	N/A	N/A	N/A
Alpha		0.05	0.05	0.05	0.05	0.05	0.05
Power		0.8	0.8	0.8	0.8	0.8	0.8
One-sided or two-sided?		Two-sided	Two-sided	Two-sided	Two-sided	Two-sided	Two-sided
Average cluster size		23	3	22	6	20	4
Number of schools	Intervention	60	60	51	51	46	46
	Control	60	60	50	50	43	43
	Total:	120	120	101	101	89	89
Number of pupils	Intervention	1380	180	1122	306	1002	281
	Control	1380	180	1100	300	943	261
	Total:	2760	360	2266	618	1945	542

Notes**.** As no pre-test data was ultimately available, the pre-test/post-test correlations as analysed report the multiple correlation between the covariates included in the analysis model and the primary outcome measure. Average cluster sizes are harmonic means of cluster sizes, which is more conservative than using the arithmetic mean in the presence of unequal cluster sizes (Bulus et al., 2019).

Attrition

Five treatment and seven control schools declined to participate in testing at the end of the intervention, and a further 146 pupils in treatment schools and 291 pupils in control schools were absent at the point of testing (despite return 'mop-up testing' in schools where three more pupils were not available on the day of main testing). Ultimately, 1002 pupils in 46 schools allocated to treatment and 943 pupils in 43 schools allocated to control are analysed. This led to an attrition rate of 20.7 percent for the intervention group and 33.1 percent for the control group, which amounted to 27.3 percent of the total randomised sample (see Table 7).

Table 7: Pupil level attrition from the trial (primary outcome)

		Intervention	Control	Total
Number of pupils	Randomised	1264	1410	2674
	Analysed	1002	943	1945
Pupil attrition (from randomisation to analysis)	Number	262	467	729
	Percentage	20.7	33.1	27.3

Pupil and school characteristics

Table 8 presents the baseline characteristics of treatment and control schools and pupils as randomised. In general, it shows that treatment and control schools are broadly similar to each other and similar to the national average for a range of characteristics. There are, however, some differences. Control schools were more likely to be located in urban settings than either intervention schools or the national average (98 percent of control schools, compared to 89 percent of intervention schools and 87 percent of all schools nationally). Control schools also differed from intervention schools along other dimensions. For example, control schools were more likely to be rated 'Outstanding' by Ofsted (33 percent versus 18 percent of intervention schools) and more likely to be academies than intervention schools (22 percent versus 16 percent).

In terms of pupil characteristics, control schools had somewhat higher average KS1 performance (15.91 for control and 15.79 for treatment schools). Again, control schools were closer to the national average (15.9). Differences in other pupil characteristics between the two groups of schools were small. Intervention schools had a slightly lower proportion of ever FSM pupils (28 percent) vs. control schools (30 percent), which was also slightly lower than the national average (31 percent). The standardised difference between treatment and control groups for this measure is 0.044 (Imbens & Rubin, 2015).

Table 8: Baseline characteristics of groups as randomised

School-level (categorical)	National-level percentage	Intervention group		Control group		
		N (missing)	%	*N* (missing)	%	
Setting: Urban	87.3	40 (5)	88.8	45 (4)	98.0	
Setting: Rural	12.7	6 (5)	11.2	1 (4)	2.0	

Ofsted: Outstanding	17.1	8 (1)	17.5	14 (0)	33.4	
Ofsted: Good	69.4	37 (1)	71.8	36 (0)	66.6	
Ofsted: RI/Inadequate	13.4	5 (1)	10.7	0 (0)	0.0	
School type: Academy	23.6	9 (0)	16.0	12 (0)	22.4	
School type: Community	41.2	25 (0)	52.0	20 (0)	48.1	
School type: Other	35.2	17 (0)	32.0	17 (0)	28.9	
School-level (continuous)	National-level mean	*N* (missing)	Mean (SD)	*N* (missing)	Mean (SD)	Standardised difference
KS1 average performance	15.9	47 (4)	15.79 (1.08)	47 (3)	15.91 (0.91)	–0.121
Pupil-level (categorical)	National-level percentage	*n* (missing)	Count (%)	*n* (missing)	Count (%)	
FSM	30.9	359 (0)	28.4	410 (26)	29.6	
Non-FSM	69.1	905 (0)	71.6	947 (26)	70.4	
EAL	15.3	474 (0)	37.5	489 (26)	35.3	
Non-EAL	84.7	790 (0)	62.5	895 (26)	64.7	

Note. School-level imbalance is calculated by applying weights for the size of the school, so that schools that are relatively more important in the pupil-level impact estimation are afforded the same importance in understanding imbalance.

Outcomes and analysis

Primary analysis

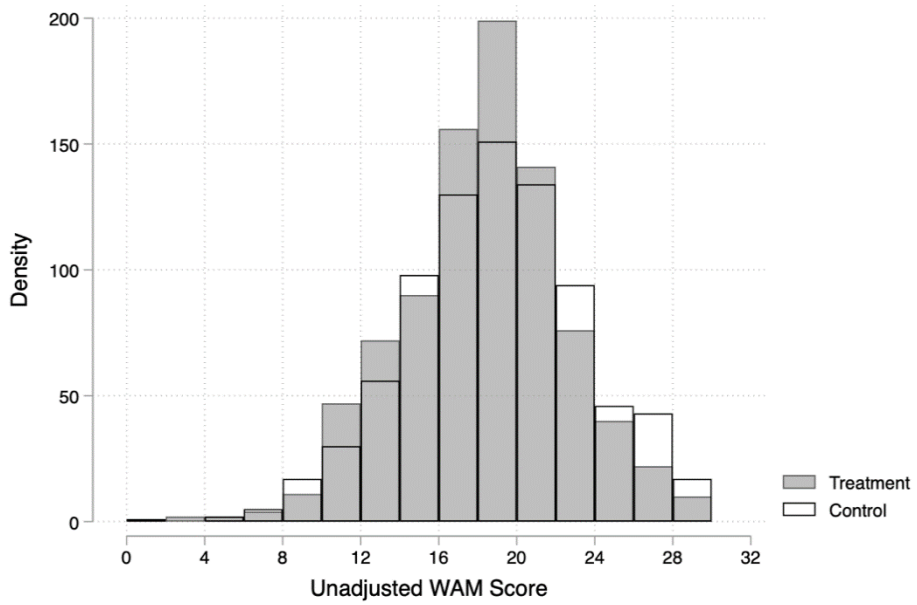
Table 9 presents the results of the analysis for the primary outcome measure. It shows the unadjusted mean for the PoP intervention group (18.42) and the unadjusted mean for the control group (18.16). After adjusting for covariates in the analysis model, we find an adjusted mean difference of 0.42. Based on this, we calculate a Hedges' *g* effect size of 0.09, which is positive, but small in magnitude (equivalent to one month's progress). It was not found to be a statistically significant difference between the two groups ($p = 0.27$).

The same unadjusted WAM scores for the intervention and the control group may be seen in Figure 4. The overall mean for the WAM across both treatment arms is 18.29 and the median is 18, which is reflected in the normal distribution of this outcome measure.

Table 9: Impact estimates

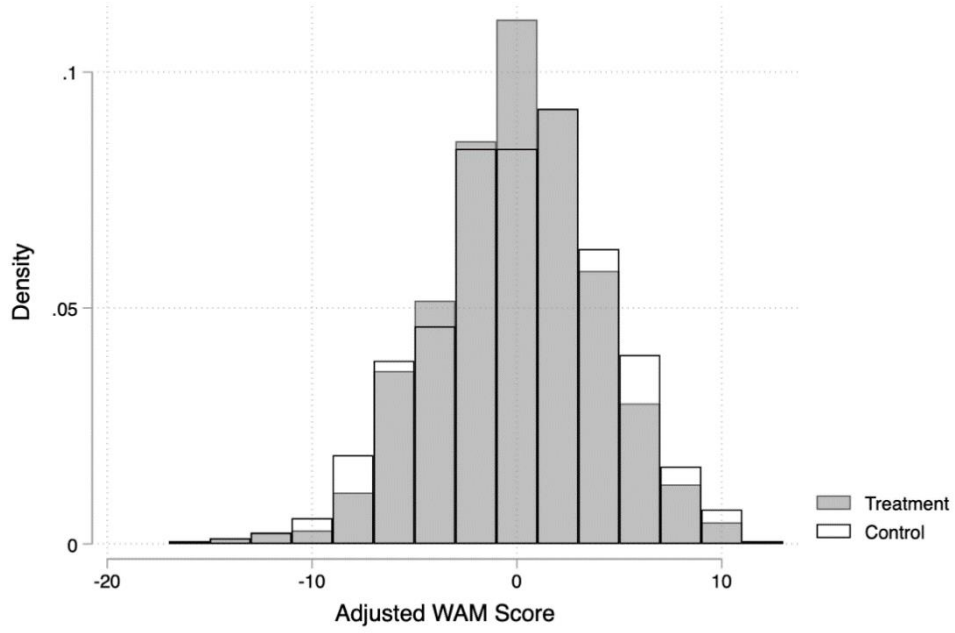
	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	*n* (missing)	Mean (95% CI)	*n* (missing)	Mean (95% CI)	Total *n* (intervention; control)	Hedges' *g* (95% CI)	*p*-value
Primary outcome							
WAM score (ideas scale double weighted)	1002 (262)	18.42 (17.81; 19.03)	943 (467)	18.16 (17.54; 18.78)	1945 (1002; 943)	0.09 (–0.07; 0.24)	0.27
Secondary outcomes							
WSEM score	983 (281)	64.83 (63.96; 65.70)	924 (486)	63.36 (62.25; 64.48)	1907 (983; 924)	0.11 (–0.00; 0.22)	0.06
Ideation score	983 (281)	20.06 (19.77; 20.35)	924 (486)	19.63 (19.24; 20.01)	1907 (983; 924)	0.09 (–0.03; 0.21)	0.14

Figure 4: Histogram of unadjusted WAM scores by treatment arm



We also present the impact analysis results for the primary outcome measure graphically through a second histogram by treatment arm displayed in Figure 5. The adjusted WAM scores for this plot are obtained from regressing the WAM writing score on pre-test measures (planned to be phonics score but, due to data access issues, instead FSM and EAL status plus class composition of these) and the variables used for stratification (i.e., the analysis model other than the treatment indicator).

Figure 5: Histogram of adjusted WAM scores by treatment arm



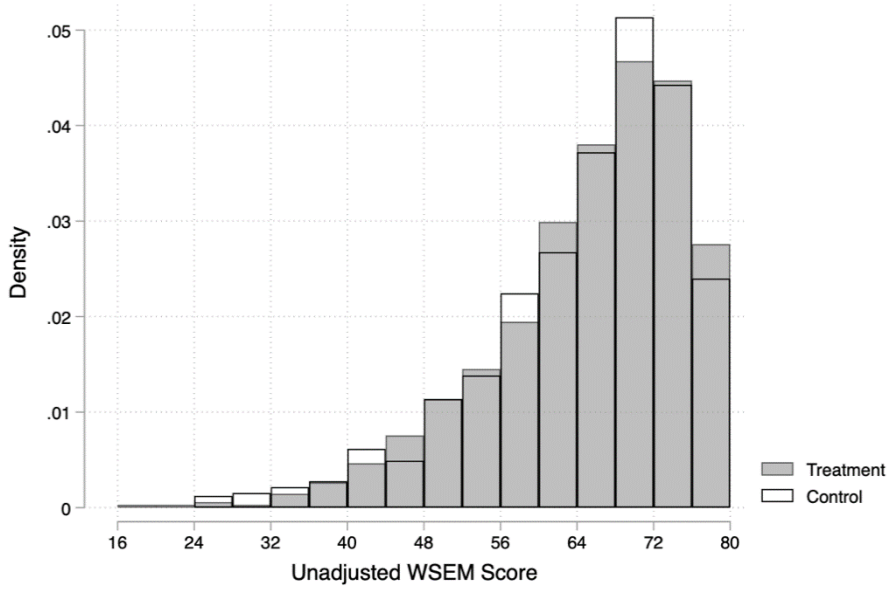
Unsurprisingly, conditioning the WAM scores on the variables used for stratification does not significantly alter the distribution for the treatment and control arms of this trial. The small mean difference between the two arms in our main analysis is evident in the plot.

Secondary analysis

The results of the secondary analysis are also presented in Table 9. The secondary outcome measures for this trial are writing self-efficacy (WSEM) and ideation. These outcomes are more closely aligned with the content and aims of the PoP intervention and so one might expect them to be more likely to register a larger effect.

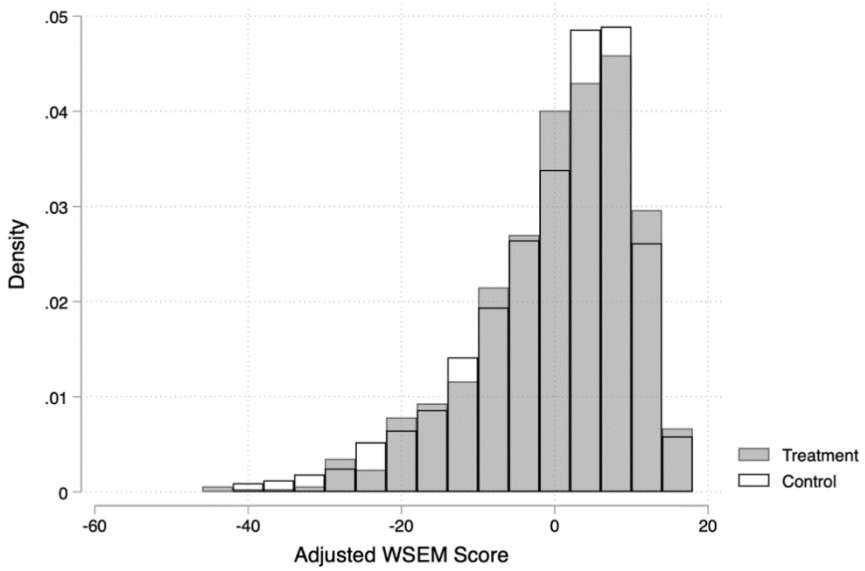
For the WSEM, there is a difference in the unadjusted means between the intervention group (64.83) and the control group (63.36), which is significant at the 10 percent significance level ($p = 0.06$). Although we generally seek statistical significance at the five percent significance level (i.e., $p < 0.05$), this positive difference provides some suggestion that the PoP intervention may have an effect on pupils' writing self-efficacy. Figure 6 shows this difference in the WSEM between the two groups graphically. It shows a right skewed distribution, indicating a high proportion of pupils giving positive responses on how they view themselves as writers, introducing some risk of ceiling effects in the analysis of this outcome measure which could attenuate the impact estimate.

Figure 6: Histogram of unadjusted WSEM scores by treatment arm



After adjusting for pre-intervention covariates the adjusted difference in means is 1.21, which translates into an effect size of 0.11. Based on EEF guidance, this effect size is equivalent to two months of progress, but is not statistically significant at the five percent significance level (p-value is 0.06). This may be seen in Figure 7, which shows the adjusted histograms of the WSEM scores by treatment arm.

Figure 7: Histogram of adjusted WSEM scores by treatment arm



A similar result emerges for the ideation score. The intervention group has a higher mean ideation score (20.06 for the intervention group and 19.63 for the control group (see Table 9 for further details and Figures E4 and E5 in Appendix E), which yields an effect size of 0.09. This effect size is not statistically significant (*p*-value is 0.14). Taken together, these positive differences, one of which is significant at the 10 percent significance level, provide some indication that the PoP intervention may have a positive impact on the secondary outcome measures.

Analysis in the presence of non-compliance

In order to examine the issue of non-compliance, we estimate the CACE. There are 11 intervention schools (22%) that did not comply based on the compliance criteria outlined in this report. Given the proportion of schools that were deemed to be non-compliant with the intervention, and that non-compliant implementation may be diluting an underlying treatment effect in compliant schools, we want to account for this lack of fidelity in calculating the effect size for the primary outcome measure.

To calculate the CACE, we use the 'ivregress' functionality of Stata to make necessary adjustments to standard errors (which are also clustered at school level) due to the instrumental variables approach (further details of this approach are available in the syntax file in Appendix F). As compliance data are available for the full primary analysis sample, this analysis is also carried out for this full sample (*n* = 1945).

The *p*-value of the exogeneity test (0.78) suggested that there would be little difference in findings between the ITT and CACE analyses. Indeed, the results of the complier analysis produced an effect size of 0.11 (95% CI: −0.08; 0.30), which is qualitatively similar to that of the ITT analysis; furthermore, the *p*-value of the CACE treatment estimate (0.27) shows that this effect size is not statistically significant. As such, we interpret these findings as not providing evidence of differential treatment effects among schools with higher levels of compliance for the primary outcome measure.

Missing data analysis

In the SAP for this trial, we outlined a missing data strategy. We noted that this strategy would be implemented if more than 5% of data in the model is missing or if more than 10% of data for a single school is missing. Due to issues with testing, a substantial number of pupils in treatment and control schools do not have outcome data. In order to test whether or not this data is MAR, we ran a logistic regression model to predict missingness in outcome data using all variables in the analysis model plus eligibility for FSM (and proportion eligible for FSM in the school), and EAL status (and proportion EAL in the school).

The results of this analysis showed that the treatment indicator and EAL were statistically significant predictors of missing outcome data (the treatment indicator is only statistically significant at the 10 percent significance level). This implies the potential for bias to complete case analysis ignoring these factors since EAL pupils might have better outcome data, which we would be more likely not to observe due to missingness. However, since all of these variables are already

included in the (revised) primary analysis model, there are no further actions possible in relation to systematic predictors of missing outcome data and our existing model should address concerns about systematic missing outcome data associated with these factors (i.e., we believe, based on our analysis, that conditional on these variables, outcome data are MAR). If any additional predictors of missingness had been discovered in this analysis, we would have run an additional ITT model controlling for them as well.

In terms of missing predictor data, none of our analysis sample is missing these data (0.0% for EAL and 0.0% for FSM). Since no one in our sample is missing this information, there is no reason to undertake multiple imputation.

Sub-group analyses

As is standard in all EEF-funded evaluations, we considered whether there is evidence of differential effects among pupils eligible for FSM as a separate sub-group (n = 542). We started by considering an augmented version of the primary analysis model, including an interaction term between the treatment variable and membership of the FSM sub-group. The estimate on this interaction is small and not statistically significant (p = 0.82), providing little evidence of a differential effect among the FSM sub-group. (See Table D2 in Appendix D for full details.)

Additional analyses and robustness checks

As outlined in the SAP for this trial, no additional analyses were planned for this trial. However, in light of ongoing discussions about appropriate interpretation of classical statistical inference, we carried out alternative statistical inference using randomisation inference to provide useful information on the extent to which there is variation between these different approaches to statistical inference. Randomisation inference is a method of conducting statistical inference using the uncertainty inherent in the randomisation process regarding the assignment of units in the trial to the treatment arms, rather than any appeal to an external sample and sampling variation (i.e., focusing on internal rather than external validity, see Cunningham, 2021). These comparative findings are reported in Table D4 in Appendix D. They result in very similar p-values despite their differing conceptual underpinnings.

Due to delays in testing some schools, we ran an additional robustness check controlling for the date of testing, since this could have affected the dosage of the intervention. The results obtained from this analysis were not substantially different from the overall impact evaluation results (i.e., the effect size is still positive but equivalent to roughly zero months progress).

We also ran a model accounting for differences in the markers who marked the WAM (the primary outcome measure), since it is possible that they introduced a new source of bias due to different degrees of leniency in marking. The inclusion of these marker fixed effects slightly attenuated the effect size, but it is still small in magnitude and positive.

Due to the imbalance in school-level KS1 scores, we also ran a model which included average school-level KS1 scores. Again, this analysis also did not substantially affect the overall effect size calculated. Imbalance in school-level Ofsted ratings was also discussed above. As such, it would seem a natural extension to run an analogous model, including this as a covariate in our analysis model. Unfortunately, due to COVID-19 restrictions, we are unable to link this characteristic to our pupil-level analysis data and, as such, are not able to undertake such an analysis. However, we note relevant points that may help to provide some guidance on what we would expect from such an analysis. Ofsted ratings are generally stronger among schools in the control group and, as such, to the extent that we might expect a positive correlation between Ofsted ratings and pupil attainment, we would anticipate this to, if anything, increase the magnitude of the positive estimated treatment effect. However, we know that Ofsted rating is not a strong predictor of pupil attainment (von Stumm et al., 2021), particularly when one has already conditioned on prior attainment, as we have done in the analysis with KS1 scores above, and, as such, would not expect this to affect our treatment estimate dramatically.

Estimation of effect sizes

As previously outlined, effect sizes are calculated using Hedges' g. These are presented in Table 9 for the primary and secondary outcomes. Appendix Table C1 in Appendix C contains additional information used in the estimation of effect sizes (e.g., the standard deviations).

Estimation of intra-cluster correlation (ICC)

As this is a clustered trial, we estimated the ICC at the class level. At the time of design, the ICC was estimated to be 0.15, which we based on EEF guidance drawing on analysis of NPD data from across England. However, in the event,

the ICC of the primary outcome measure (the WAM) has turned out to be considerably higher (at 0.31). There are aspects of the sample which may explain some or all of this difference, including that the schools participating in the trial are disproportionately drawn from urban areas (although this was also the case for the estimate using EEF guidance, since we used the highest regional ICC, which happened to be for London), and that our missing data analysis suggests the potential for our sample to differ in terms of EAL composition. That said, we doubt these explain the extent of the difference, which is likely attributable to the outcome measure itself, for example reflecting within-class shared writing practices that are consistently captured by the measure. This should be borne in mind in the design of future trials making use of measures of this type.

Implementation and process evaluation (IPE) results

Summary

- The majority of teachers attended all training sessions and case study data showed they found them to be useful opportunities for seeing how pupils would experience the PoP intervention.
- In case studies, teachers highlighted the creative and flexible nature of the programme, which allowed them to embed PoP into their teaching of the national curriculum.
- They also highlighted the intensive nature of the intervention and the difficulty of timetabling 20 lessons for the intervention, but noted that having the end goal of producing a picturebook created momentum for the intervention.
- Case study data raised concerns about a shortage of art supplies and their associated costs, which proved challenging for some schools.
- Survey data and case studies highlighted the high degree of senior leadership support for the intervention, with members of SLT attending CPD sessions and PoP lessons, which was viewed as an important factor for supporting the success of the intervention.
- Case study teachers perceived the intervention as having a positive impact on pupils, particularly those who face challenges with traditional literacy instruction.
- Teachers also expressed their changed beliefs about the positive value of picturebooks and the possibility for incorporating aspects of the intervention into future teaching practice.

This section of the report contains the findings of the IPE, which are based on data collected from case studies, training observations, and the supplementary post-intervention survey (see the *Methods* section for more details). The subsections are structured by theme rather than research question, in line with what Nowell et al. (2017) propose as best practice qualitative reporting; however, the research questions addressed in each subsection are outlined in Table 10 below. The first subsection provides background information on the case study schools and comparisons with usual practice. The second subsection addresses the factors influencing programme delivery, and the third subsection describes pupil and teacher engagement with the PoP programme. The final subsection details the mechanisms underpinning the intervention and its perceived impact. Along with information gathered through the case studies, this section draws on administrative data and findings from the survey of intervention schools.

Table 10: Implementation and process evaluation (IPE) findings and research questions

Section in IPE findings	Research questions
Context	RQ5. How does a pupil's home learning environment and exposure to books affect their engagement?
Implementation	RQ1.In what ways was the programme implemented? What are the barriers and facilitators of delivery? In particular: SLT 'buy-in'; delivery of training and resources: (a) whether it appears to be effective in ensuring that teaching staff understand the aims and main features of the intervention [see '*Not Answerable*' for part (b)]; delivery of the intervention: (a) whether it appears to facilitate children's engagement [see '*Not Answerable*' for part (b)].
	RQ2.To what extent did the schools engage with the intervention in line with the intervention aims?

	RQ3. How was the quality of the intervention perceived by teachers, senior leaders and teaching assistants?
	RQ7. How did teachers experience and perceive the PoP training?
Engagement	RQ8. How did pupils engage with the author-illustrator workshop element of the programme?
Perceived mechanisms and impact	RQ6. How does teachers' response to creative writing change as a result of taking part in the training? How is this reflected in teacher confidence in writing and drawing, and in classroom practice in PoP lessons and more generally?
	RQ9. Are there any unforeseen mechanisms in the intervention that appear to be bringing about change?
	RQ1(c). Delivery of the intervention – whether it appears to be effective in supporting children's attainment.

Context and motivations for participation

Desire to improve pupil outcomes

There were three main motivations identified for case study schools who applied to be part of the PoP trial. The first is related to wanting to improve pupil outcomes. Interviewees, especially members of the SLT, explained that they wanted to take part in order to improve reading comprehension and English standards at their schools. Specifically, interviewees felt that PoP could support children who are disadvantaged or have limited access to books at home, who they identified as experiencing particular difficulties with reading comprehension. Only 40% of treatment survey respondents felt that their pupils had access to a range of books at home, indicating that this may have been a motivation to take part among schools in the trial.

> *'[Power of Pictures] really stood out as being quite an engaging way for those children that are disadvantaged... They don't enjoy reading. They don't enjoy writing... It was really appealing that using picturebooks to improve comprehension, and improve their understanding and writing could really make a difference.' (SLT member 01)*

Interest in working with CLPE

A second motivating factor influencing some schools was the reputation of the Centre for Literacy in Primary Education (CLPE). Interviewees, both teachers and SLT members, described having either having had direct experience with CLPE, for instance, through taking part in their Power of Reading programme, or having heard about it from a colleague. They believed that CLPE was a respected organisation and positive previous experiences made them excited about taking part in PoP.

> *'[I knew of CLPE before coming to this school, and then I came here and gradually got more involved with the literacy. So, I went on lots of courses there, and I myself did Power of Reading… We wanted our curriculum to be based on research and pedagogy, and I feel that's what CLPE can offer for us.' (SLT member 06)*

Different approach from usual practice

A final motivating factor was that PoP offered the opportunity to try a new approach to teaching English. SLT interviewees said that it was an opportunity to upskill their teachers to be able to teach using picturebooks, which they felt were a useful teaching tool for Key Stage 2. Some teachers initially had concerns about how they would plan lessons involving

picturebooks for older children. They wondered whether Year 5 classes would find picturebooks childish, and felt nervous about the prospect of teaching art and drawing. However, overall teachers were enthusiastic about taking part. Teachers reported that the usual approach to teaching literacy is highly structured, typically involving pupils reading longer novels, and writing with a beginning, middle and end structure, with a strong emphasis on spelling and grammar. In comparison, they felt PoP would offer a different, more creative approach to teaching literacy.

> *'I liked it because I would say by the end of summer one and summer two [the final two teaching half terms in an academic year], I felt that my teaching got a bit stale, and I was repeating cycles of things that I had done before. I wanted fresh stuff.' (Teacher 02)*

> *'I was very excited... I haven't heard of this project before, so that was a bit of an unknown entity for me, and there was also a bit of trepidation because of reasons that I've already said, not being particularly gifted as an artist as well. But overall yes, pretty happy to be involved with it.' (Teacher 05)*

Survey data indicated that teaching using picturebooks was not usual practice for the majority of control schools. In the last year, 58% of control schools reported that they had only used picturebooks a 'few times', with 17% reporting that they 'never' utilised picturebooks to teach (see Appendix L for full response rate). Case study interviews with teachers revealed that one motivation for some schools getting involved in the programme was so that they could introduce the use of picturebooks. Some case study SLT members and teachers felt that the use of picturebooks in teaching had been overlooked due to the focus placed on writing and reading attainment in the curriculum.

> *' I think some of our picturebooks had dropped, I think at one period there was a drive towards more books that have lots of text, because it was felt that it would develop their writing.' (SLT Member 06)*

Implementation fidelity

Delivery compliance and fidelity

According to the PoP model, delivery with fidelity involved teachers delivering ten PoP lessons during the autumn and spring terms (meaning a total of 20 lessons). The compliance criterion required that at least half of these were delivered. Administrative data showed that by the end of the school year, 32 intervention schools (63%) had completed all 20 lessons, 11 schools (22%) had completed 15 lessons, and eight schools (16%) had completed 10 or fewer. In terms of the case study schools, three out of six case study teachers had not delivered all 20 lessons to their classes at the point of sampling. These teachers described delivering the sessions over a longer period of time or delivering the content over a greater number of lessons. The reason given by teachers who delivered more than the intended 10 lessons per term was because the intervention was new to both them and the pupils, delivering more lessons helped pupils to become familiar and comfortable with the different approaches used in the programme.

> *'I just think that it didn't fit into ten sessions discreetly because of how long it took the children to become – not engaged, but less fearful of putting their ideas on the page.' (Teacher 05)*

The reasons given for extending the timeframe in which the intervention was delivered was related both to teachers needing to familiarise and become comfortable with the intervention, and because they felt it enabled them to deliver the programme at the pace that was appropriate for their pupils.

Programme delivery also differed from what was prescribed in the PoP manual in terms of the type of picturebook used by teachers. Teachers explained that they preferred to not always use PoP-specific materials because they thought materials online or books by other authors would be more engaging for pupils. Some teachers also said that they did not always adhere strictly to the lesson sequence that was recommended by PoP. One of the reasons given for this was that the lesson sequences contained too much detail for the length of time they had been assigned to deliver their literacy sessions, causing teachers to select some elements of the lesson plan and leave out other parts.

> *'So, the scheme that I was given I didn't particularly stick to it in a strict manner. There were a couple of points that I liked. I know what they are like as a class, and I know that stuff is right up their alley, and they will do beautifully.' (Teacher 02)*

In contrast to schools where delivery of PoP varied from what was intended, schools that were able to deliver the programme with fidelity explained that this was because PoP did not differ very much from their school's own approach to literacy.

> *'So Power of Pictures has worked really well with that, because the talk and discussion around pictures books really fits with how we teach literacy here.' (SLT member 01)*

Factors influencing fidelity of implementation

There were five factors highlighted by teachers that affected delivering the programme as intended (see *Intervention* section for detail): teacher training; the unstructured and creative nature of the programme; timetabling; availability of resources; and senior leadership support. These are discussed in turn below.

Teacher training

As part of the PoP programme, teachers were required to attend two and a half days of training. This was also one of the compliance criteria. Overall, 44 (86%) intervention class teachers, including all of the case study teachers, attended all of the training days. During these sessions, teachers learnt about the craft of picturebook making by working directly with an author-illustrator and a member of the PoP delivery team. The aim of the training was to equip teachers with the necessary skills and knowledge required to deliver PoP in the classroom.

Teachers reported that the training sessions served to strengthen their understanding of PoP. The survey found that 95% of respondents (41 of 43 surveyed) felt confident that they understood the main features of the intervention, with 63% (27 of 43 surveyed) stating that they were 'very confident'. Observations of the training sessions indicated that trainers detailed the different phases of the project and provided teachers with an in-depth explanation of the benefits of picturebooks. Teachers who attended the training reported that they valued being informed about how PoP differed from what they perceived to be the standard English curriculum, further contributing to their understanding of the intervention. Additionally, teachers said that the sessions provided them with greater insight into pupil experiences of the programme, explaining that this was largely due to the modelling component of the training, which provided teachers with first-hand experience of the activities,

> *'If we as adults hadn't been asked to do the task... I wouldn't have realised how difficult it was and had that empathy for the children.'* (Teacher 05)

The survey indicated that 98% of responding teachers (42 of 43) felt prepared to deliver PoP in the classroom after attending CPD sessions (with 76% reporting that they felt 'very prepared'). Interviews with teachers indicated that one strength of the training sessions was that the sessions adopted a practical approach, providing teachers with exercises and techniques that they could easily translate to the classroom. The author-illustrator also spent time in the sessions talking teachers through the creative process, which helped teachers who did not have an artistic background to understand how to begin to approach the programme, and in turn how to support their pupils to engage with the intervention. Teachers also explained that PoP lesson plans provided and discussed during the training enabled them to understand the natural progression of lessons and understand how to organise these sessions. This was particularly the case in the context of teachers having limited time for lesson planning.

> *'Once you are back at school, and you have got a thousand things to juggle, it was so useful having that laid out for you and not having to start from scratch with each lesson but having that ready.'* (Teacher 06)

While the practical, interactive training was highly valued, there was a comment made that the training could be content heavy in places.

> *'[The] training sometimes can get a bit laborious and sit down and get lectured at.'* (SLT member 02)

Teachers also noted that the CPD sessions provided them with space to reflect and discuss the practicalities of delivering PoP in the classroom with one another. This was described as one of the most valued aspects of the sessions.

> *'Having a chance to look at those activities and kind of do them with the other teachers and then really reflect on, "actually what literacy skills are we teaching through this?" That's the kind of thing that helped me most' (Teacher 06)*

The unstructured and creative nature of the programme

The unstructured and creative nature of the programme supported delivery for two main reasons. Firstly, teachers said that the programme was less restrictive for pupils with lower levels of attainment compared to the national curriculum, because the PoP approach was perceived to place less emphasis on grammar, punctuation and spelling at the point of ideation and creation. Teachers thought that this helped to 'level the playing field' between pupils with higher and lower levels of attainment. Similarly, teachers said that the use of picturebooks enabled pupils with EAL to understand and immerse themselves in lessons in a way they would not normally be able to. Secondly, children enjoyed having the freedom to create their own characters and stories without having to worry about formal 'rules' of language, such as syntax.

> *'Breaking the rules, so being able to, you know, you get people saying that you can't start a sentence with a conjunction, and it's able to have a bit of freedom with those. So breaking the rules has been quite liberating I think for some of the children.'* (SLT member 03)

Conversely, some pupils struggled with the level of freedom inherent in the programme. Specific areas of the intervention that pupils struggled with included the freedom associated with 'rough work' and also the idea that their drawings did not have to be perfect. Teachers said that certain pupils required guidance around what they were expected to write and what their drawing should look like, in order to feel confident and comfortable with the exercises. For instance, some pupils who were highly self-critical about their drawing skills were asked to find photos to imitate, rather than drawing from their heads. In particular, teachers reported that pupils with autistic spectrum conditions could struggle with the PoP approach and often required additional support from teachers.

> *'I think the blank page at the start was the main thing... the children were more like, what do you want my page to look like?' (SLT member 05)*

> *'He just totally lacked confidence when it came to draw[ing], and he was saying Mr [teacher's name], can you draw it for me. So he really was refusing to engage with it.'* (SLT member 04)

Some teachers recommended that it would be helpful for CLPE to provide more guidance during the training sessions about how to manage the freedom given to pupils in the classroom.

Timetabling

Another barrier to implementation emphasised by teachers was the challenges associated with timetabling. Teachers said that national curriculum requirements made it difficult for them to allocate time to deliver the required 20 sessions, with the core curriculum often being prioritised over the delivery of PoP sessions. This was a particular issue for teachers if they felt they did not have direct support from their SLT in delivering the programme. Therefore, some teachers recommended that CLPE provide information and guidance to school SLTs about how teachers can deliver the intervention, while also meeting national curriculum requirements; for example, how to deliver PoP, including technical aspects such as spelling, punctuation and grammar.

Availability of resources

The need for schools to provide children with art resources could also present difficulties. The programme recommends that schools provide pupils with materials such as colouring pencils, watercolour paints, acrylic paints, collage materials, tracing paper and publication paper. However, teachers explained that their school did not have a separate budget for the programme and instead used their English or Literacy budgets, which could not always stretch to this.

> *'It's really hard, because I'm in charge of English, and I have a budget that is basically gone.'* (SLT member 03)

The perception that the programme was resource-intensive put teachers under increased strain, with some teachers reporting that they had spent their own money to buy resources for the programme. In other cases, schools used their literacy budget to buy materials for the programme, which meant that other school requirements could not be funded,

though what these were was not specified. Similarly, some schools were not able to afford to buy any additional resources for the programme and instead used resources that the school already had.

> *'He just used general class resources, so anything we had. We have no money to buy anything, so whatever was there.'* (SLT member 02)

> *' There's lots of little things that we just don't have at the moment. Things like visualisers, something to put their work on screen more rapidly. We've got a few ageing cameras, or we can just go and scan it somewhere on the printer, but that's all inconvenient.'* (Teacher 02)

Senior management support

Another key issue was the degree of SLT 'buy-in'. In some schools, the SLT had active involvement in the programme; for instance, by creating opportunities for teachers to share learning about the programme with other teachers, or by attending PoP lessons and CPD sessions themselves. This gave them first-hand experience of what the intervention entailed, as well as insight into pupil experience of the intervention. In some schools, teachers said that SLT enthusiasm for the programme had been central to keeping the programme going in the school.

In other cases, the SLT had not been directly involved in the programme. However, in some of these cases, teachers still felt that SLT members had been on hand to provide them with practical support and advice, particularly with navigating problems, such as timetabling issues and arranging cover so that teachers could attend the training.

> *'Because I'm the senior leader, my aspect has been... for her to talk it through with me. We've talked about how she can implement it, what she needs from us as a school. So, there has been resourcing implications and things like that so helping her with those sorts of practical aspects.'* (SLT member 03)

Teachers said that support from the teacher responsible for leading the English curriculum in their school, in particular, had been key. Though English leads themselves are not always members of school SLTs, teachers felt that discussions with English leads gave them a link through which changes or resources required to deliver PoP could be communicated with the SLT.

Engagement

This section explores pupil engagement with the PoP programme, and specifically pupil engagement with the author-illustrator workshop, which had an 100% attendance rate among intervention classes.

Teachers reported that pupils had a positive experience at the workshop and were particularly excited by the opportunity to meet the author-illustrator, whose book they had been reading.

> *'The children treated him like a superstar, because he had illustrated and written a book. He was famous to them because they had read the book before they met him.'* (Teacher 05)

Teachers felt that the author-illustrator's presence at the workshop was key to engaging pupils and that this experience would not have been as 'eye-opening' for students if they had not met the author-illustrator. Teachers also said that the down to earth approach adopted by author-illustrators allowed students to feel comfortable enough at the workshop to ask questions.

Survey data showed that 93% of respondents (40 of 43 surveyed) felt that the author-illustrator workshop motivated pupils to produce their own picturebooks, with 72% (31 of 43 surveyed) saying that they 'strongly agreed'. Teachers explained that this was due to the author-illustrators showing pupils simple steps that they could take to create their characters, which made the story writing and drawing process less daunting. Pupils described enjoying learning about the 'secrets' of writing and illustrating, such as the concept of using five basic shapes to create characters, a technique used by all of the author-illustrators taking part in the evaluation. Pupils also said that the author-illustrator workshop taught them how to begin to think of ideas for their stories, for example by looking at magazine and newspaper clippings. In addition, teachers said that the author-illustrator workshops were key to pupils overcoming their worries around creating perfect and neat characters, due to author-illustrators emphasising that rough work is part of the process.

> *'I think it was really nice for them to hear about the writing process and the illustration process from an author and illustrator. It opened their eyes a little bit about the resilience to that [process] and their attitudes to writing. We do one redraft, and children can struggle to redraft writing sometimes because it's done. But then [the author-illustrator] showed them a pile of paper that is just scribbled on, and it's that sort of not being neat.'* (Teacher 03)

Teachers said that the workshop encouraged pupils to take ownership of their characters, which in turn provided them with the confidence to create the characters that they wanted to. However, some teachers suggested that the structure of the sessions should be adapted so that pupils do not spend prolonged periods sitting and listening to the author-illustrator, instead suggesting that the sessions should be split up to include more breaks.

Mechanisms and perceived impact

This section explores the perceived impact of the intervention on pupils, teachers or the wider school, and the mechanisms believed to underpin the impact.

Pupil mechanisms and perceived impact

Overall, teachers felt the programme had had a positive impact on pupils' writing, with 84% of survey respondents (36 of 43 surveyed) indicating that the programme was perceived to have had a positive impact on pupils' writing skills (23% stating the impact was 'very positive', 10 of 43 surveyed). Teachers said improvements had been seen in the additional detail pupils had included in their descriptions and stories. Teachers thought that this was caused by the fact that using pictures encouraged students to think more deeply and carefully about their writing.

> *'The level of their thinking for a picturebook has deepened their thinking because they have got to think more because there isn't as many words.'* (SLT member 03)

Teachers described how some pupils' writing had also improved in relation to other literacy lessons. They explained that pupils had a better awareness of the writing process, which had contributed to strengthening their sense of narrative structure. This included having a better understanding of the purpose of each element in a story structure and enabling them to produce better structured, more coherent stories. Additionally, teachers noted that pupils' interest in writing had increased, with some actively seeking out opportunities to write outside of their PoP sessions.

A new outcome has been added to the logic model to represent the greater level of interest pupils developed in picturebooks. Teachers reported that some pupils were more interested in picturebooks both in and outside the classroom, with interviewed pupils also describing how they had learnt that pictures can communicate as much about what is happening in a story as words can. Other interviewed pupils explained that the images in picturebooks helped them to understand writing better.

A perceived impact of the programme was its effect on pupils' confidence, with 95% of teacher respondents (41 of 43 surveyed) reporting that the intervention had a positive impact (of which 53% reported that this impact was 'very positive', 23 of 43 surveyed). Teachers felt this was the result of the ownership pupils had over their stories.

> *'The girl that you saw who read her story today, her confidence has grown massively. There is no way at the beginning of this year that she would have stood up with her story at the front. But she took those pictures and she was really keen to share that because it was drawing, and she knew she was good at it.'* (Teacher 05)

The freedom associated with the programme was also felt to help remove initial reservations some pupils had around drawing and writing, which was reinforced by the idea that there is not one correct way to write a story or draw a picture. Teachers suggested that the reduced focus on spelling and punctuation at the point of ideation and creation also encouraged reluctant writers, and gave them self-belief to write their stories.

The focus of drawing in PoP led some pupils to spend more time drawing at home. Prior to their participation in the programme, pupils had had limited opportunities to draw in class, thus teachers said that the emphasis on drawing led to some pupils enjoying drawing more even if they were not skilled at it.

> *'So actually, a lot of the drawing focus that we had in the project, there was a huge development more in attitude than in drawing skills.'* (Teacher 06)

The drawing focus was seen as especially beneficial for pupils who were not confident or skilled writers, enabling them to communicate through pictures.

In terms of creativity, 98% of survey respondents (42 of 43 surveyed) reported a perceived positive impact in pupils' creative thinking as a result of taking part in PoP, with 67% reporting a 'very positive' impact (29 of 43 surveyed). Teachers felt that the ideation component of the programme had developed pupils' ability to link visual imagery and text to tell a broader story. They also explained that some students had become more imaginative in their writing and were willing to take more risks when creating stories.

> *'If I think about kids like [name], he's a bit [unimaginative]. Now he is less [unimaginative]. Before he would stick to simply tried and tested things that he had seen elsewhere that is going to produce a sensible piece of writing and do that.'* (Teacher 02)

Teachers linked this with the freedom associated with PoP, giving pupils the confidence to move away from the norm and permitting them to try something they are less familiar or comfortable with.

The data from case study interviews also indicated that teachers felt the programme had encouraged pupils to develop additional skills outside of reading and writing. Most notably, the programme was seen as having had a positive impact on the planning skills of pupils with low attainment and/or SEN. Teachers said that the PoP process provided pupils with time to develop their ideas and opportunity to plan their narrative structure. Teachers also explained that the sense of ownership associated with PoP had permitted pupils to plan in the way they felt most comfortable, and begin their stories at the points at which they felt most comfortable, an experience they did not normally have in the classroom.

> *'I have one little boy who really struggles to generate ideas and struggles with coming up with his own ideas. When he did the picturebook, he just started with the blob and this blob suddenly became a story about the water cycle, so it became a raindrop. He had a story about the water cycle and now he's confident.'* (Teacher 03)

Some teachers also said that the programme supported pupils with lower levels of attainment to develop resilience. They felt that the process led to students having a clearer understanding of the steps involved in creating their picturebooks, thereby decreasing the likelihood of pupils giving up part of the way through, and motivating them to complete their stories.

While the programme was seen as having a positive effect on pupils with lower levels of attainment or with SEN, some teachers thought that the programme had less of an impact on pupils with higher levels of attainment in writing, because these teachers felt that these pupils might feel restricted by needing to use pictures and writing less lengthy pieces of work.

Mechanisms and perceived impact on teachers

Some teachers said that the programme had a positive impact on their own confidence, including teachers who previously felt that they could not draw who reported developing more confidence in their drawing abilities over the course of the programme. This, in turn, led to teachers feeling confident in their capability to support pupils who struggled to draw, since they felt they could empathise with their experience.

> *'I never was good at it [art] at school. Suddenly, we were doing painting and doing cat stories, and if I can do that, all the kids could do something and sort of enjoy it.'* (Teacher 01)

Teachers said that their confidence in drawing was encouraged by attending CPD days and working with the author-illustrator. In particular, the modelling and scaffolding techniques used during the day were often cited as simplifying the drawing process, which helped to foster confidence in teachers.

Similarly, there were teachers who said that the programme encouraged them to step out of their comfort zone. For example, some teachers who were initially embarrassed to share their work with others became more comfortable doing so, and attributed this to the training, which gave them the opportunity to meet other teachers with similar concerns. They explain that this supported them to feel less awkward about their work and confident enough to draw in class.

Participation in the programme also impacted how some teachers approach teaching. These teachers felt that their previous approach was less engaging than it could have been.

> *'I would say I was probably dull in my approach to that. This was back in summer too and I was skint on ideas. It was a really enjoyable book and I probably felt like the kids didn't enjoy it as much as they would have just because of my approach to the book.'* (Teacher 02)

Teachers described moving away from more traditional approaches, both in literacy and in other subjects that they taught. Some teachers reported incorporating elements of PoP such as ideation and the additional editing stage into their literacy classes, as they valued the impact that these elements had had on pupils. In addition, teachers felt that the programme made them aware of how important it was to explain things explicitly to pupils and adopted the use of modelling across a range of subjects, including subjects not directly related to writing such as maths.

Teachers also described that the programme had changed some teachers' perceptions of picturebooks. These teachers explained that they had previously avoided using picturebooks in KS2, due to the absence of words. However, PoP gave them a better understanding of how images could be used to convey meaning and information.

> *'Yeah, like I said, we have always kind of done it [teach picturebooks], but this has just given me such a better understanding of how to do it, and strategies for doing it so it's richer and there is more breadth to it.'* (Teacher 06)

> *'So it really opened my mind to that actually and that's the message I brought back to school, because it's not just about looking at the nice pictures but then reading the actual words. It's about conveying meaning in really very subtle ways and the kind of inference they need to make from a picture is just as complex as it is for inferring something from words.'* (Teacher 06)

Wider perceived impact

Participating teachers reported sharing the key principles of the programme and their experience of delivering PoP with other teachers. In some schools, this led to a growing interest among SLT members and teachers in spreading the programme more widely in the school. A number of teachers said that their schools felt passionate about not losing the perceived positive impact of the programme and would like to expand it across all year groups.

> *'I think that's the beauty of it. I think it can be adapted to any year group.'* (SLT member 03)

As a result, in some schools teachers who delivered the programme during the trial were tasked with training teachers in different year groups, to embed learning and begin rolling out the programme. However, it was felt that funding was a key barrier to the wider expansion of the programme. Schools said that they have a limited budget and had to consider other areas that required funding before a budget can be allocated to PoP. Teachers in these schools recognise that PoP would benefit pupils and mentioned that they would seek to identify funding for the programme in the coming years.

Some schools actively engaged parents during the process and kept them informed of their children's progress and dedicated PoP events for parents, to showcase the work that pupils had done, were organised.

> *'We had a parent celebration at the end which was lovely. We had all the children with their planning books and books on the table, and the parents came in and looked at the children's books.'* (Teacher 03)

Cost

Delivery of the PoP intervention cost approximately £352 per school for the year it was delivered, and the majority of costs were realised in the first year. Materials and printing costs included purchasing books and art supplies. As described above, some schools had the necessary supplies and did not need to purchase additional materials, while others needed to either purchase a number of resources or implemented PoP without them due to lack of available budget. The range and average costs are presented in Table 11. Expenses were incurred in relation to travel and subsistence to attend training and transport pupils to the author-illustrator workshop. Programme fees were paid to the developer.

To calculate the total cost per pupil over three years, we assumed the number of pupils would cumulatively increase from 25 pupils in year one, 50 in year two and 75 in year three. We assumed 25 pupils as this was the number of treated pupils. Based on these assumptions, the total cost per pupil per year over three years is £5 (set out in Table 11). The cost breakdown is set out in Table 12 (note that all figures are rounded to the nearest pound).

Table 11: Cost of delivering Power of Pictures (PoP)

Item	Type of cost	Average cost (£) (Min, Max)	Total cost over 3 years (£)	Total cost per pupil per year over 3 years (£)
Materials and printing	Ongoing	27 (£0, £100)	80	1
Expenses	First Year	123 (£0, £460)	123	2
Programme fees	First Year	200 (£200, £200)	200	3
Physical environment	First Year	3 (£0, £12)	3	<1
Total		352	406	5

Source: Cost case study interviews with teachers (*n* = 4).

Table 12: Cumulative costs of Power of Pictures (PoP) (assuming delivery over three years)

	Year 1	Year 2	Year 3
Power of Pictures	£351	£379	£406

Source: Cost case study interviews with teachers (*n* = 4).

Training

To support the delivery of the intervention, schools allocated one teacher to attend 2.5 days of training between October 2018 and March 2019. Of the case study schools, three attended the recommended number of training sessions. However, one of the case study schools had only attended one of the training sessions. In terms of staff time, training was estimated at 20 hours per school. Assuming that the same teacher is supporting the intervention for three years, the teacher would only have to attend the training in the first year. As per EEF cost evaluation guidance, teacher supply cover is not included as part of programme costs.

Preparation

There was some additional administration time associated with supporting the intervention. Teachers reported that in the first year they spent approximately three hours per school preparing for the programme (ranging from two to five hours over the course of the year). Assuming the same level of preparation is required each year, staff time for preparation over three years would amount to approximately nine hours.

Delivery

The recommended number of sessions required for the delivery of the intervention is 20 sessions across two terms. On average, schools delivered the programme for 17 hours or two days per school per year. The reported hours of delivery ranged from 10–20 hours per year. In addition, as part of the intervention teachers were required to take pupils to one author-illustrator workshop session. Of the case study schools, only one missed this session and on average teachers

spent four hours or 0.5 days per school accompanying pupils to visit author-illustrators. The reported hours required for the author visits ranged from four to seven hours per year.

Conclusion

Table 13: Key conclusions

Key conclusions
1. Pupils who received the Power of Pictures (PoP) programme had, on average, higher writing scores (equivalent to one month of additional progress) as compared to children in the control group. This is our best estimate of impact which has a moderate to high security rating. However, as with any study, there is uncertainty around the result: the possible impact of this programme on reading attainment ranges from one month less progress to positive effects of three additional months of progress.
2. Among children eligible for free school meals (FSM), those in schools that PoP was delivered in also made one additional month's progress. These results may have a lower security rating than the overall findings because of the smaller number of pupils in this group.
3. Children in PoP schools had higher writing self-efficacy and writing creativity (ideation) scores than those from schools in which the programme was not taught. These differences suggest that PoP may have a positive impact on these outcomes.
4. The visual element of this programme attracted learners who traditionally have difficulties engaging in literacy activities.
5. Teachers reported high levels of engagement with the programme, not only from the pupils and themselves, but also from the senior leadership team (SLT) at their schools. This said, implementing PoP was perceived to be very time intensive.

Impact evaluation and IPE integration

The goal of this impact evaluation was to answer three key questions:

1. What is the effect of participating in PoP over the course of one school year on pupils' writing skills?
2. What is the effect of participating in PoP over the course of one school year on pupils' writing self-efficacy?
3. Does participating in PoP over the course of one school year have an impact on pupils' perception of their own capacity to generate ideas?

The purpose of the IPE was to understand how teachers experienced the PoP training, how teachers delivered PoP lessons in their classroom, and, in particular, to better understand the barriers and facilitators to implementation and delivering the intervention with fidelity. The IPE was also used to identify how the teaching through PoP compares to usual practice, and to determine the cost to schools of delivering the intervention.

Evidence to support the logic model

The original PoP logic model posited that the impact of the intervention on writing outcomes may have an effect through pupils' engagement with and motivation for writing, which may in turn have an effect on children's sense of efficacy as a writer. This was well supported by the IPE data and somewhat confirmed by the impact evaluation data.

The IPE survey data indicated that teachers were very positive in terms of the perceived effect of the intervention on pupils, saying that improvements had been seen in the additional detail pupils had included in their descriptions and stories. Teachers thought that this was caused by the fact that using pictures encouraged students to think more deeply and carefully about their writing. They also highlighted increased interest in writing, especially from pupils who were seen as non-traditional learners (e.g., pupils who had difficulty engaging with the traditional writing curriculum or did not view their writing skills positively). This was somewhat reflected in the impact evaluation findings, which showed positive, but statistically insignificant, effects on the WSEM and the ideation score. The results for the WSEM were equivalent to two months' progress, but only statistically significant at the 10 percent significance level. The effect size for the ideation score was also positive and equivalent to one month's progress, but again, not statistically significant.

Building on the IPE analysis, a new outcome has been added to the logic model to represent the greater level of interest pupils developed in picturebooks. Teachers reported that some pupils were more interested in picturebooks both in and outside the classroom, with pupils also describing how they had learnt that pictures can communicate as much about

what is happening in a story as words can. Other pupils explained that the images in picturebooks helped them to understand writing better.

Interpretation

The results of this trial found a small, positive overall effect on the WAM (the primary outcome), providing suggestive evidence of a positive effect even though it is not statistically significant. This effect size is equivalent to one month's progress. These results were limited by power issues arising from data access issues and a higher than anticipated ICC, which makes it harder to detect a statistically significant impact.

The impact evaluation results on the secondary outcome measures, the WSEM and the ideation score, were somewhat larger and also positive, but again not statistically significant at traditional significance levels. Taken together this provides suggestive evidence that the PoP intervention had a positive impact on pupils' non-cognitive outcomes.

The PoP intervention was delivered by an experienced delivery partner. CLPE has a strong track record in delivering CPD. The survey data showed that teachers were very engaged in the PoP intervention. The survey responses to the IPE surveys show a high level of buy-in, not only from the participating teachers, but also from the SLT. At some schools SLT members attended training sessions in order to show their support for the programme. Teachers reported feeling like they were learning alongside their pupils.

Some teachers reported spending significantly longer on this intervention than had been planned (four weeks) because they had a clear goal: producing a picturebook. Other teachers reported that they found it challenging to complete the intervention. This was highlighted in the compliance data, where 11 treatment schools were deemed to be non-compliant (i.e., they did not complete the required number of lessons). However, the overall impact evaluation results were not different from the compliance results.

The PoP programme was previously evaluated with a report released in 2017. That process evaluation was highly favourable, finding that the 'courses [were] a very impressive success' (Horner & White, 2017). The evaluators found the programme to have a lasting effect both on pupils (as perceived by their teachers) and the author-illustrators who participated. They specifically highlighted the intervention's role in using creativity to prepare for and develop writing. The findings from this evaluation show suggestive evidence of the programme's positive effect on writing skills, writing self-efficacy and ideation, which are not as strong as the findings from the previous evaluation.

Limitations and lessons learned

Baseline balance checks revealed some imbalance across trial arms with, on average, higher KS1 scores and more positive Ofsted ratings at the schools in the control group compared to the intervention group (particularly in the analysis sample, as opposed to the randomised sample, implying this could be induced by systematic attrition, discussed further below). The intervention schools also had higher proportions of EAL pupils. We acknowledge the potential for this to lead to bias in our point estimates (although, to the extent that it is due to chance in the randomisation, this is part of the uncertainty in estimates captured by the confidence intervals). Other things being equal, higher KS1 scores in the control group would imply our estimate to be an underestimate of the true impact (we have also carried our robustness checking on this point); the same is true for the higher Ofsted rating in the control group (we were not able to carry out analogous robustness checking, although correlation between KS1 scores and Ofsted ratings and relatively weak predictive power of Ofsted ratings for pupil attainment over and above this relationship with KS1 scores leads us to doubt this would reveal substantial bias). The imbalance in EAL is also likely to be mitigated by the inclusion of this factor as a covariate in our primary analysis model.

Unfortunately, due to issues in accessing the NPD, planned pre-test measures (planned to be phonics screening check scores) were not available at point of analysis. This reduced the explanatory power of outcome measure from covariates included in the primary analysis model compared to our expectations at design stage (although, of course, these were assumptions and so could still have turned out to be too optimistic, even if we did have phonics screening check scores available). This, combined with greater attrition than assumed and a higher intra-cluster correlation of the outcome measure than was anticipated (based on our assumptions and previous EEF guidance), led to a substantial increase in the MDES between the design stage and the analysis stage. This increases the uncertainty around our impact estimates and, hence, decreases the likelihood of finding a statistically significant impact of the intervention for a given effect size. The higher than anticipated ICC of the outcome measure (which we judge to be unlikely to be attributable to the

composition of our sample of schools) suggests the need for further preparatory work in understanding this aspect of non-NPD outcome measures ahead of trial design – previous literature on the outcome measure did not cover school-level clustering and neither could our small-scale piloting.

Robust assessment of writing is challenging, particularly during primary schooling. There are few measures available, and none have been used in a similar context. The measure used in this trial, the WAM (Dunsmuir et al., 2015), was a pragmatic choice, which comes with some limitations (e.g., in terms of how relatively new it is); however, existing evidence suggests that it is a valid, consistent and reliable measure. Nevertheless, despite the many dimensions of writing it aims to capture, it may not have been able to capture those proximately affected by this intervention. This does present a limitation of this evaluation, albeit one we feel we have done all we could to address, given the measures available.

Following randomisation, five intervention and seven control schools were lost to follow-up. This led to substantial attrition for this trial. We have to acknowledge that attrition carries with it a risk of bias (Sterne et al., 2019) and one that cannot typically be remedied analytically. In short, those schools and pupils with missing outcomes data could have affected the result from the trial in different ways – meaning that the result is biased away from the 'true' estimate of impact.

Additionally, of the 51 schools allocated to the intervention arm, 11 were deemed to be non-compliant with the treatment (i.e., the teacher did not attend training sessions and deliver at least 50% of the lessons and the class did not attend the workshop). The compliance analysis for this trial indicated that accounting for this did not change the overall impact evaluation results. As previously mentioned, many of the schools that were compliant spent much more time receiving the intervention, since teachers allocated more time than the four weeks planned to the creation of the picturebooks.

It was also more difficult to conduct outcome data collection than had been anticipated, resulting in differences in when schools were tested. A robustness check including the date of testing did not reveal any substantial differences in the impact evaluation results.

Generally, the data collected from intervention schools as part of the IPE (either via survey or fieldwork visits) only represent the views and experiences of a subset of the larger treatment population. The qualitative findings are therefore not statistically representative, though the use of purposive sampling means that they should provide a good indication of the range and diversity of experiences and attitudes. Additionally, there may be some recall errors in survey responses.

Similarly, the case study approach to the cost evaluation represents the range and diversity of costs encountered in implementation among highly engaged schools. Sampling was done to capture variation of spend among schools with high/low proportions of FSM pupils and per pupil spending, as these were hypothesised to correlate with costs. It is possible that the sampled schools were not representative of typical costs of full implementation of the programme. However, the programme fee was the largest cost to schools and did not vary by school. Spending ranges were provided for other direct costs, so that prospective schools could consider costs they may encounter above and beyond paying for the programme itself.

Future research and publications

This study leaves a number of issues relating to the impact of PoP unaddressed. First, it has not been able to provide evidence of the impact of PoP on medium-term nationally collected literacy outcomes, such as KS2 national curriculum tests, which would help reinforce its instrumental value to primary schools. Second, given its particular aims around improving engagement and motivation with writing, it may be of particular value to explore variation in the impact of the approach depending on pupils' pre-existing motivation (perhaps coarsely proxied by prior attainment). We also think there would be value in exploring potential spillovers between PoP and activity and attainment in other parts of the curriculum. These are concrete areas where we see the possibility for future research.

An additional overarching report on all five Learning about Culture interventions funded by EEF and the RSA will be published in 2021. This will include the three Key Stage 2 Learning about Culture interventions (Craft of Writing, Power of Pictures and the Young Journalist Academy) and the two Key Stage 1 Learning about Culture interventions (First Thing Music and Speech Bubbles). This report will pool outcome data across the trials for a combined impact evaluation and synthesise IPE results across all interventions.

In addition, we plan to publish several academic journal articles summarising the findings from this and the other trials in the Learning about Culture project, as well as discussing their implications for future research on arts education learning interventions and also methodological issues including the challenges of measuring children's writing skills at scale.

References

Angrist, J.D., Pischke, J.-S., 2008. *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Bloom, H.S., Richburg-Hayes, L., Black, A.R., 2007. Using covariates to improve precision for studies that randomize schools to evaluate educational interventions. *Educational Evaluation and Policy Analysis* 29, 30–59. https://doi.org/10.3102/0162373707299550

Bruning, R., Dempsey, M., Kauffman, D.F., McKim, C., Zumbrunn, S., 2013. Examining dimensions of self-efficacy for writing. *Journal of Educational Psychology* 105, 25.

Bulus, M., Dong, N., Kelcey, B., Spybrook, J., 2019. PowerUpR: Power analysis tools for multilevel randomized experiments. R package version 1.0.4. https://CRAN.R-project.org/package=PowerUpR

Catterall, J., 2009. *Doing well and doing good by doing art: The effects of education in the visual and performing arts on the achievements and values of young*. Los Angeles/London: Imagination Group/IGroup Books.

Catterall, J.S., 2012. *The arts and achievement in at-risk youth: Findings from four longitudinal studies*. Research Report# 55. National Endowment for the Arts.

Cunningham, S., 2021. *Causal inference: The mixtape*. Yale University Press.

DfE, 2016*. Schools, pupils and their characteristics: January 2016*, SFR20/2016. Department for Education, London.

DfE, 2017. *Progress in International Reading Literacy Study (PIRLS): National Report for England*. Department for Education, London.

Doonan, J., 1993. *Looking at pictures in picturebooks*. Thimble Press.

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.

EEF, 2015. *Intra-cluster correlation coefficients*. Education Endowment Foundation, London

EEF, 2018. *Statistical analysis guidance for EEF evaluations*. Education Endowment Foundation, London.

Gerber, A.S., Green, D.P., 2012. *Field experiments: Design, analysis, and interpretation*. WW Norton.

Graham, S., Perin, D., 2007. A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology* 99, 445.

Gutman, L., Schoon, I., 2013. *The impact of non-cognitive skills on outcomes for young people: Literature review*. Education Endowment Foundation, London.

Hannes, K., 2011. Chapter 4: Critical appraisal of qualitative research. In J. Noyes, A. Booth, K. Hannes, A. Harden, J. Harris, S. Lewin, C. Lockwood (Eds*.), Supplementary guidance for inclusion of qualitative research in Cochrane systematic reviews of interventions (version 1)*.

Hedges, L.V., 1981. Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics* 6, 107–128. https://doi.org/10.3102/10769986006002107

Horner, S., White, J., 2017. Power of Pictures: Final Project Evaluation. Centre for Literacy in Primary Education (CLPE).

Ilie, S., Sutherland, A., Vignoles, A., 2017. Revisiting free school meal eligibility as a proxy for pupil socio-economic deprivation. *British Educational Research Journal* 43, 253–274. https://doi.org/10.1002/berj.3260

Imbens, G.W., Rubin, D.B., 2015. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Kiefer, B.Z., 1995. *The potential of picturebooks: From visual literacy to aesthetic understanding*. Merrill.

Murphy, V.A., Kyriacou, M., Menon, P., 2013. *Profiling writing challenges in children with English as an additional language (EAL)*. University of Oxford, Department of Education 14.

Neelands, J., Belfiore, E., Firth, C., Hart, N., Perrin, L., Brock, S., Holdaway, D., Woddis, J., 2015. *Enriching Britain culture, creativity and growth: 2015 report by the Warwick Commission on the future of cultural value*. University of Warwick, Coventry.

Newman, M., Bird, K., Tripney, J., Kalra, N., Kwan, I., Bangpan, M., Vigurs, C., 2010. *Understanding the impact of engagement in culture and sport: A systematic review of the learning impacts for young people. CASE: The culture and sport evidence programme*.

Nikolajeva, M., Scott, C., 2013. *How picturebooks work*. Routledge.

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J., 2017. Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods*. doi: 10.1177/1609406917733847

Pajares, F., 1996. Self-efficacy beliefs in academic settings. *Review of Educational Research* 66, 543–578.

Ritchie, J., Lewis, J., Nicholls, C.M., Ormston, R., 2013. *Qualitative research practice: A guide for social science students and researchers.* Sage.

Rubin, D.B., 2004. *Multiple imputation for nonresponse in surveys*. John Wiley & Sons.

Sterne, J.A., Savović, J., Page, M.J., Elbers, R.G., Blencowe, N.S., Boutron, I., Cates, C.J., Cheng, H.-Y., Corbett, M.S., Eldridge, S.M., 2019. RoB 2: a revised tool for assessing risk of bias in randomised trials. *The BMJ* 366.

Strand, S., Malmberg, L., Hall, J., 2015. *English as an additional language (EAL) and educational achievement in England: An analysis of the National Pupil Database*. London: Education Endowment Foundation.

Torgerson, C.J., Torgerson, D.J. 2013. *Randomised trials in education: An introductory handbook*. London: Education Endowment Foundation.

UK Government, 2019. Get information about schools. [online]. https://get-information-schools.service.gov.uk/

Appendix A: EEF cost rating

Figure A1: Cost rating

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

Appendix B: Security classification of trial findings

OUTCOME: Writing Assessment Measure (WAM)

Rating	Criteria for rating			Initial score		Adjust		Final score
	Design	**MDES**	**Attrition**					
5 🔒	Randomised design	<= 0.2	0-10%					
4 🔒	Design for comparison that considers some type of selection on unobservable characteristics (e.g. RDD, Diff-in-Diffs, Matched Diff-in-Diffs)	0.21 - 0.29	11-20%					
3 🔒	Design for comparison that considers selection on all relevant observable confounders (e.g. Matching or Regression Analysis with variables descriptive of the selection mechanism)	0.30 - 0.39	21-30%	3 🔒		Adjustment for threats to internal validity **[0]**		3 🔒
2 🔒	Design for comparison that considers selection only on some relevant confounders	0.40 - 0.49	31-40%					
1 🔒	Design for comparison that does not consider selection on any relevant confounders	0.50 - 0.59	41-50%					
0 🔒	No comparator	>=0.6	>50%					

Threats to validity	Risk rating	Comments
Threat 1: Confounding	Low	Some concerns regarding imbalance in school characteristics (e.g., Ofsted rating), discussed in the report.
Threat 2: Concurrent Interventions	Low	Limited information about concurrent interventions.
Threat 3: Experimental effects	Low	Limited evidence of experimental effects.
Threat 4: Implementation fidelity	Low	Compliance in terms of attending training sessions and delivering the planned 20 lessons was fairly high. Some evidence from case studies that implementation fidelity varied but these variations seemed to be in the spirit of PoP. Little difference between ITT and CACE analysis.
Threat 5: Missing data	Moderate	There was rather a lot of missing data, with about 27% of the randomisation sample missing. While this is a threat to validity, a padlock has already been dropped for attrition, so no further adjustments are necessary. The lack of NPD and pre-test data also causes some concerns here, though there is little more the authors could have done under the circumstances explained in the report.

Threat 6: Measurement of outcomes	Low/ Moderate	Some concerns about the primary outcome variable, which the authors address at length in the report. Piloting, administration and marking procedures reported provide adequate reassurance.
Threat 7: Selective reporting	Low	The study differs from the original protocol, but the reasons are justified to a satisfactory extent. Selective reporting not apparent.

- **Initial padlock score:** 3 Padlocks – Randomised design with 0.21 MDES at randomisation and 27.3% attrition.
- **Reason for adjustment for threats to validity**: N/A – One moderate and one low/moderate risk, with the direction of likely biases unclear. No further adjustments required.
- **Final padlock score:** initial score adjusted for threats to validity = 3 Padlocks

Appendix C: Effect size estimation

Table C1: Effect size estimation

			Intervention group		Control group			
Outcome	Unadjusted differences in means	Adjusted differences in means	n (missing)	SD of outcome	n (missing)	SD of outcome	Pooled SD	Hedges' correction (*J*)
WAM Score (ideas double weighted)	0.26	0.42	1002 (262)	4.8	943 (467)	4.73	4.77	0.99961
WSEM Score	1.46	1.21	983 (281)	10.65	924 (486)	11.52	11.08	0.99961
Ideation Score	0.43	0.34	983 (281)	3.65	924 (486)	3.82	3.73	0.99961

Note. SD refers to standard deviation.

Further appendices:

Appendix D: Additional tables

Table D1: Baseline characteristics of groups as analysed (n = pupil, N = school)

School-level (categorical)	National-level percentage	Intervention group		Control group		
		N (missing)	%	N (missing)	%	
Setting: Urban	87.3	36 (5)	90.3	38 (4)	97.0	
Setting: Rural	12.7	5 (5)	9.7	1 (4)	3.0	
Ofsted: Outstanding	17.1	7 (1)	15.9	11 (0)	25.2	
Ofsted: Good	69.4	33 (1)	72.3	32 (0)	74.8	
Ofsted: RI/Inadequate	13.4	5 (1)	11.8	0 (0)	0.0	
School type: Academy	23.6	6 (0)	11.9	11 (0)	24.1	
School type: Community	41.2	23 (0)	52.1	18 (0)	45.9	
School type: Other	35.2	17 (0)	36.0	14 (0)	30.0	
School-level (continuous)	National-level mean	N (missing)	Mean (SD)	N (missing)	Mean (SD)	Standardised difference
KS1 average performance	15.9	43 (3)	15.80 (1.09)	40 (3)	16.04 (0.91)	0.238
Pupil-level (categorical)	National-level percentage	n (missing)	%	n (missing)	%	Standardised difference
FSM	30.9	281 (0)	28.0	261 (0)	27.7	
Non-FSM	69.1	721 (0)	72.0	682 (0)	72.3	

EAL	15.3	385 (0)	38.4	287 (0)	30.4	
Non-EAL	84.7	617 (0)	61.6	656 (0)	69.6	

Note. School-level imbalance is calculated applying weights for the size of the school, so that schools which are relatively more important in the pupil-level impact estimation are afforded the same importance in understanding imbalance.

Table D2: Sub-group analyses

Model	Hedges' *g* (95% CI)	N	*p*-value of interaction term	*p*-value of treatment variable
WAM FSM sub-group analysis	0.06 (–0.20, 0.32)	542	0.82	0.67
WSEM FSM sub-group analysis	0.02 (–0.08, 0.12)	525	0.61	0.67
Ideation FSM sub-group analysis	0.14 (–0.05, 0.33)	525	0.35	0.16

Table D3: Full sample summary statistics

Outcome	Mean	SD	ICC	N
WAM Score (ideas scale double weighted)	18.29	4.77	0.31	1945
WSEM-16 Score	64.12	11.1	0.19	1907
Ideation Score	19.85	3.74	0.19	1907

Table D4: Sampling and randomisation inference p-values

Outcome	Effect size	Sampling inference *p*-value	Randomisation inference *p*-value
WAM Score (ideas scale double weighted)	0.09	0.27	0.32
WSEM Score	0.11	0.06	0.06
Ideation Score	0.09	0.14	0.14

Table D5: Complier average causal effect (CACE) analysis

Model	Hedges' *g* (95% CI)	N	First stage F test	Compliance / Treatment correlation	*p*-value of treatment variable

Compliance analysis	0.11 (–0.08, 0.30)	1945	F(8, 88) = 53	0.82	0.27

Table D6: Robustness checks

Model	Effect size (Hedges' *g*)	N	Treatment coefficient	*p*-value of treatment variable
Primary impact evaluation	0.09	1945	0.418	0.273
Test date	0.08	1932	0.395	0.303
Marker fixed effects	0.07	1945	0.344	0.369
School level KS1	0.09	1945	0.445	0.250

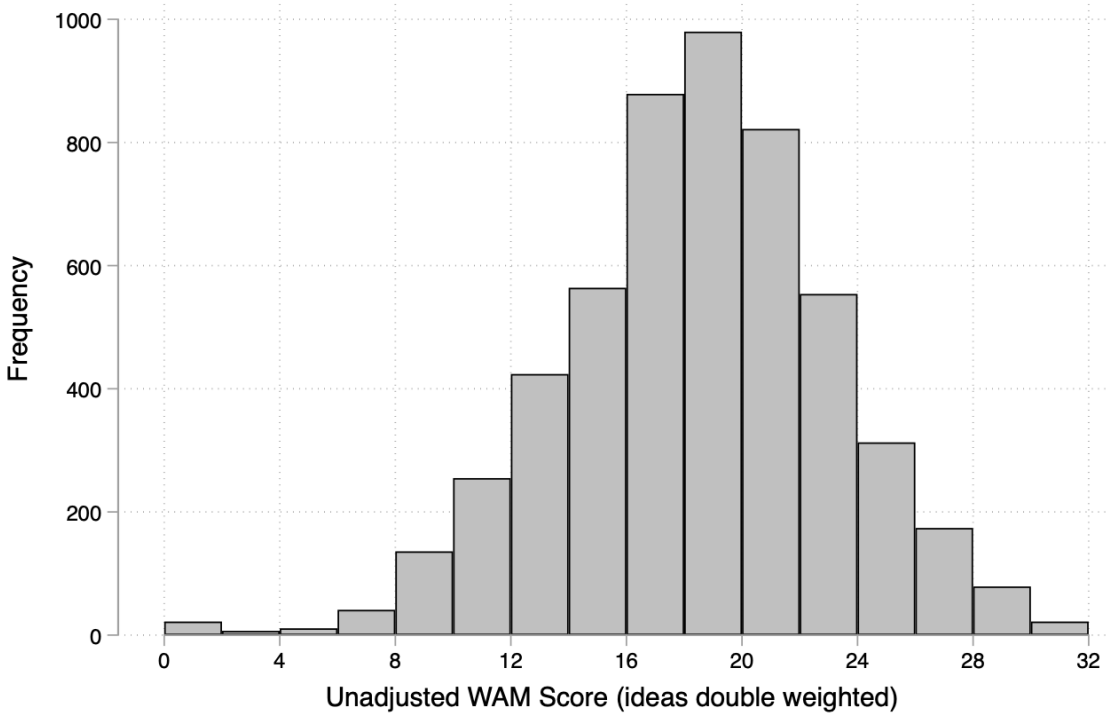
Table D7: Number of treatment and control schools (*N*) in each stratum

Stratum	Treatment (*N*)	Control (*N*)
Low EAL/Low FSM	14	13
Low EAL/High FSM	12	12
High EAL/Low FSM	12	12
High EAL/High FSM	13	13

Note. The strata are defined by the intersection of English as an additional language (EAL) students (high/low split at sample median) and proportion of Free School Meals (FSM) students (high/low split at sample median).

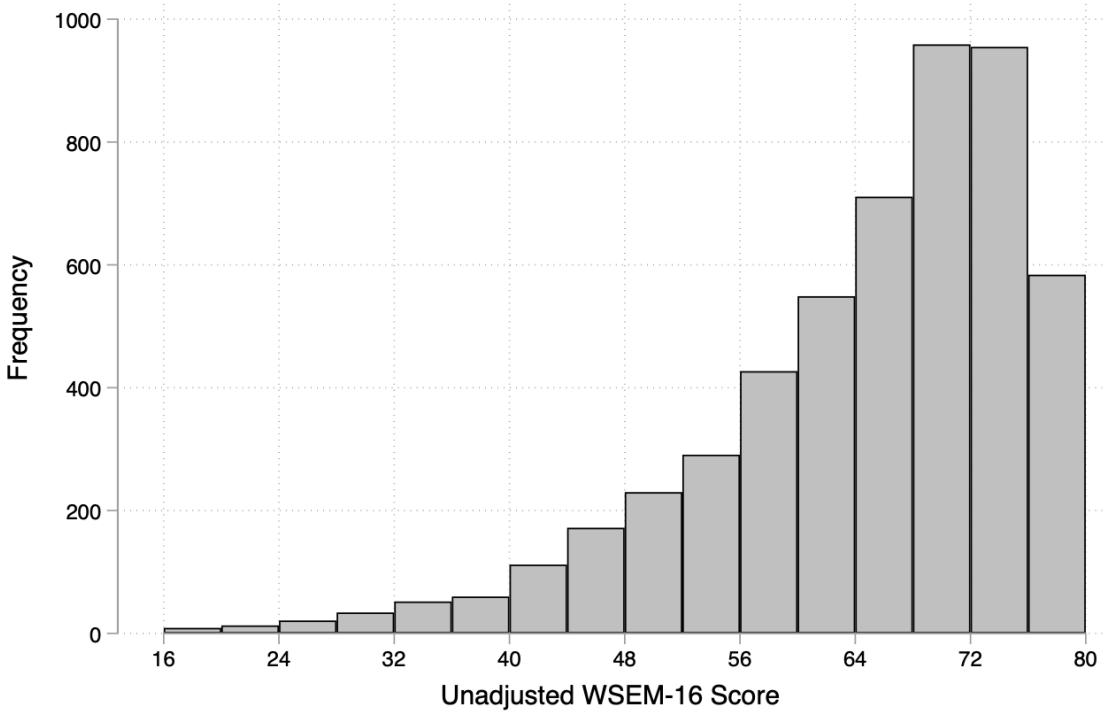
Appendix E: Additional histograms

Figure E1: Histogram of WAM scores pooled across all three KS2 Learning about Culture trials



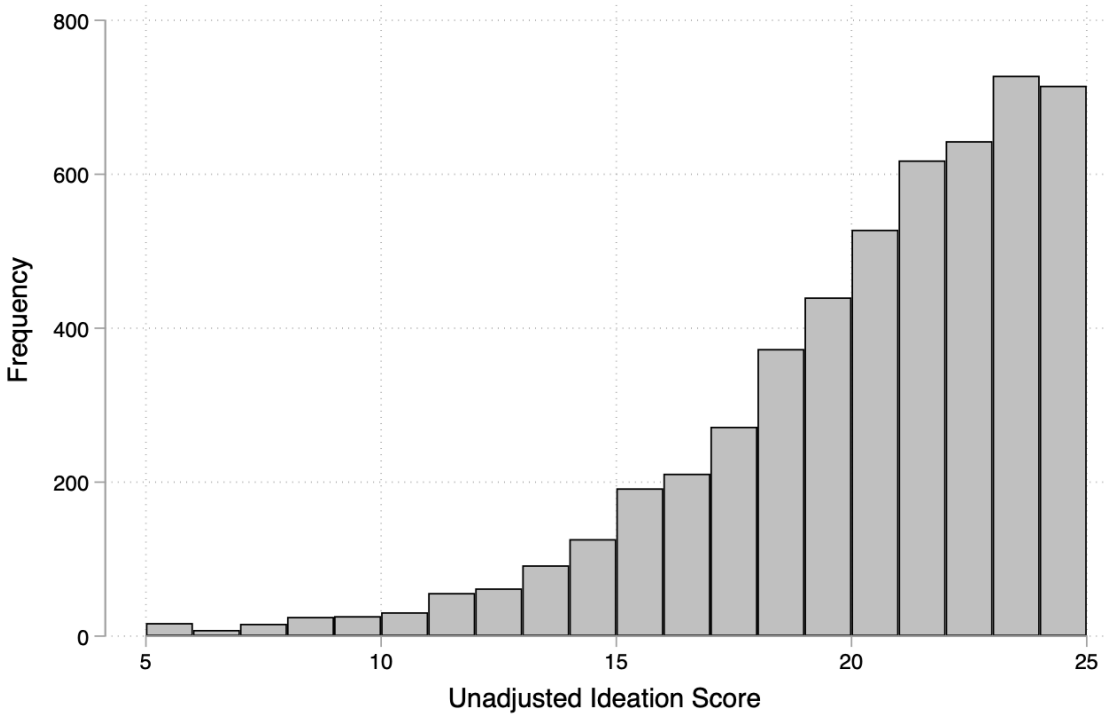
Note. This histogram contains all WAM outcome data collected for the KS2 Learning about Culture trials (Craft of Writing, Power of Pictures, and Young Journalist Academy)

Figure E2: Histogram of WSEM scores pooled across all three KS2 Learning about Culture trials



Note. This histogram contains all WAM outcome data collected for the KS2 Learning about Culture trials (Craft of Writing, Power of Pictures, and Young Journalist Academy)

Figure E3: Histogram of Ideation scores pooled across all three KS2 Learning about Culture trials



Note. This histogram contains all WAM outcome data collected for the KS2 Learning about Culture trials (Craft of Writing, Power of Pictures, and Young Journalist Academy).

Figure E4: Histogram of unadjusted ideation scores by treatment arm

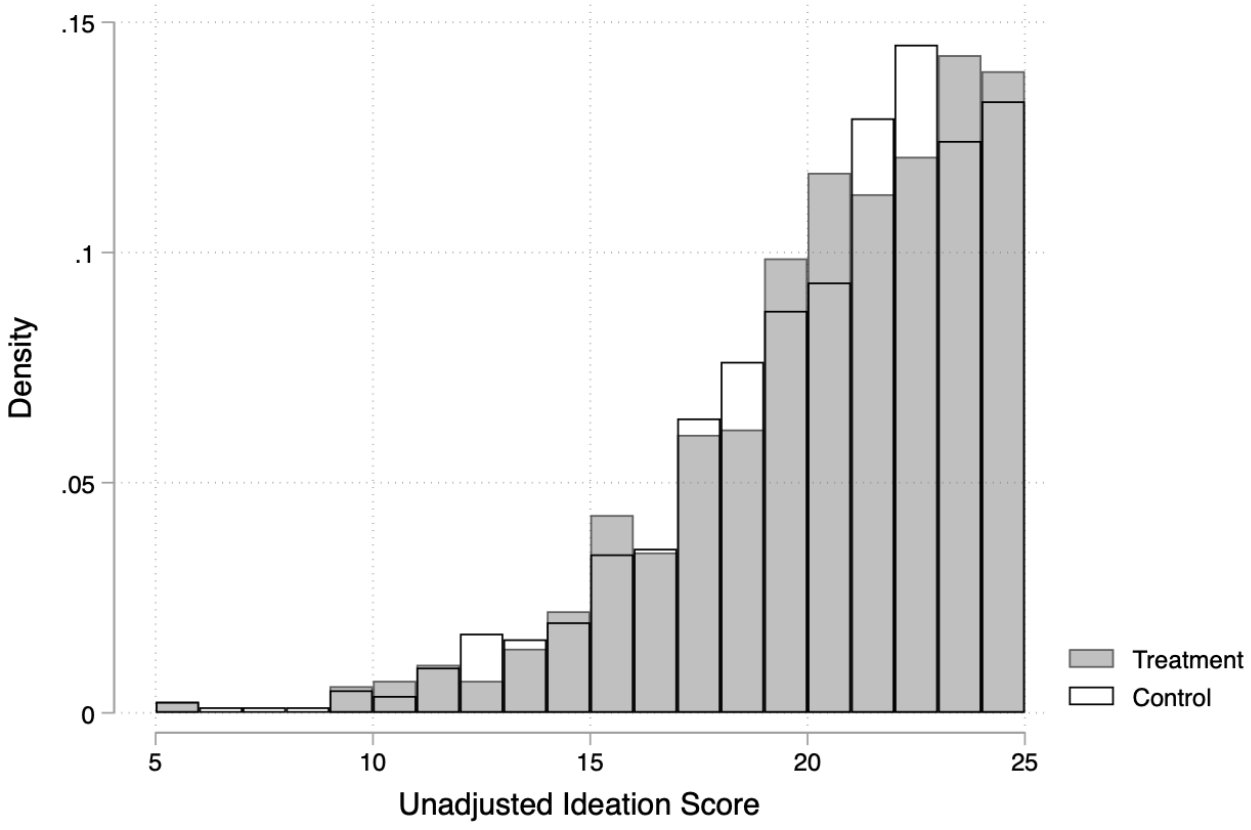
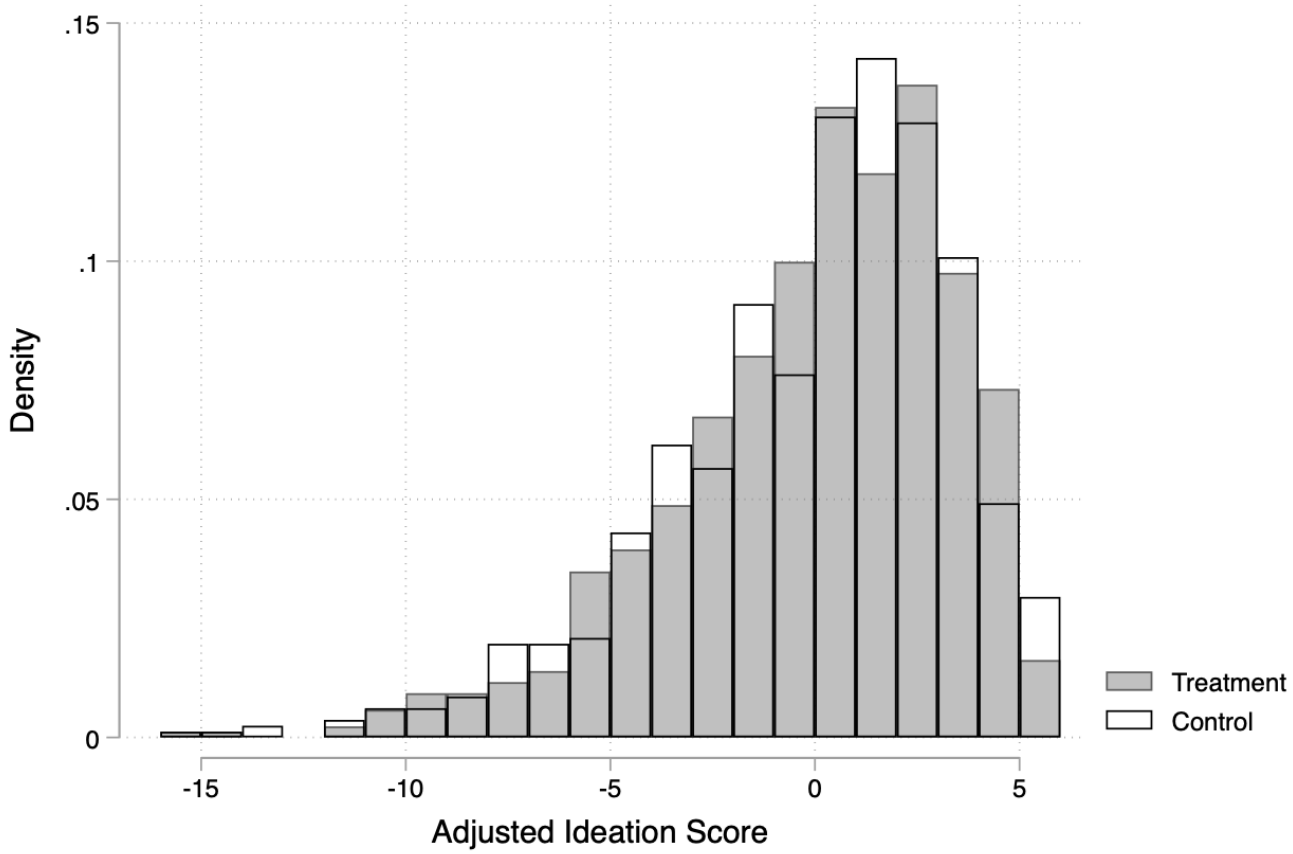


Figure E5: Histogram of adjusted ideation scores by treatment arm



Appendix F: Analysis syntax

```
ssc install blindschemes
run ‘[INSERT FILE PATH HERE]/lac_ks2_nationalmeans.do’
global workd ‘[INSERT FILE PATH FOR GLOBAL MACRO HERE]’
global datad ‘[INSERT FILE PATH FOR GLOBAL MACRO HERE]’
cd ‘${workd}’
clear

use ‘${datad}/pop_data.dta’

* Greate sample variables
cap drop primarysample
gen primarysample = 0
replace primarysample = 1 if wam_score_dw<. & FSM<. & EAL<. & block<. & anonschoolid<. & treat<.

count if primarysample == 1
local primarysample_n = r(N)
count if primarysample == 1 & treat==1
local primarysample_treat_n = r(N)
count if primarysample == 1 & treat==0
local primarysample_control_n = r(N)

count if primarysample == 0
local primarysample_miss = r(N)
count if primarysample == 0 & treat==1
local primarysample_treat_miss = r(N)
count if primarysample == 0 & treat==0
local primarysample_control_miss = r(N)

cap drop secondarysample
gen secondarysample = 0
replace secondarysample = 1 if wsem16_score<. & FSM<. & EAL<. & block<. & anonschoolid<. & treat<.

count if secondarysample == 1
local secondarysample_n = r(N)
count if secondarysample == 1 & treat==1
```

```
local secondarysample_treat_n = r(N)
count if secondarysample == 1 & treat==0
local secondarysample_control_n = r(N)

count if secondarysample == 0
local secondarysample_miss = r(N)
count if secondarysample == 0 & treat==1
local secondarysample_treat_miss = r(N)
count if secondarysample == 0 & treat==0
local secondarysample_control_miss = r(N)

cap drop randomisedsample
gen randomisedsample = 0
replace randomisedsample = 1 if block<. & anonschoolid<. & treat<.
count if randomisedsample == 1
local randomisedsample_n = r(N)
count if randomisedsample == 1 & treat==1
local randomisedsample_treat_n = r(N)
count if randomisedsample == 1 & treat==0
local randomisedsample_control_n = r(N)

local primaryattrition_n = `randomisedsample_n' – `primarysample_n'
local primaryattrition_treat_n = `randomisedsample_treat_n' – `primarysample_treat_n'
local primaryattrition_control_n = `randomisedsample_control_n' – `primarysample_control_n'
local primaryattrition_percent : di %7.1fc 100–((`primarysample_n'/`randomisedsample_n')*100)
local primaryattrition_treat_p : di %7.1fc 100–((`primarysample_treat_n'/`randomisedsample_treat_n')*100)
local primaryattrition_control_p : di %7.1fc 100–((`primarysample_control_n'/`randomisedsample_control_n')*100)

count if wam_score_dw < .
local primarymeasure_n = r(N)
count if wam_score_dw < . & treat==1
local primarymeasure_treat_n = r(N)
count if wam_score_dw < . & treat==0
local primarymeasure_control_n = r(N)

count if wam_score_dw >= .
local primarymeasure_miss = r(N)
```

```
count if wam_score_dw >= . & treat==1
local primarymeasure_treat_miss = r(N)
count if wam_score_dw >= . & treat==0
local primarymeasure_control_miss = r(N)

count if wsem16_score < .
local wsem_n = r(N)
count if wsem16_score < . & treat==1
local wsem_treat_n = r(N)
count if wsem16_score < . & treat==0
local wsem_control_n = r(N)

count if wsem16_score >= .
local wsem_miss = r(N)
count if wsem16_score >= . & treat==1
local wsem_treat_miss = r(N)
count if wsem16_score >= . & treat==0
local wsem_control_miss = r(N)

count if ideation_score < .
local idea_n = r(N)
count if ideation_score < . & treat==1
local idea_treat_n = r(N)
count if ideation_score < . & treat==0
local idea_control_n = r(N)

count if ideation_score >= .
local idea_miss = r(N)
count if ideation_score >= . & treat==1
local idea_treat_miss = r(N)
count if ideation_score >= . & treat==0
local idea_control_miss = r(N)

* School-level descriptives
quietly tab anonschoolid
local asrandom_schoolcount = r(r)
quietly tab anonschoolid if treat==0
```

```
local asrandom_schoolcount_control = r(r)
quietly tab anonschoolid if treat==1
local asrandom_schoolcount_treat = r(r)

quietly tab anonschoolid if primarysample==1
local wam_schoolcount = r(r)
quietly tab anonschoolid if secondarysample==1
local wsem_schoolcount = r(r)
local idea_schoolcount = r(r)

quietly tab anonschoolid if treat==0 & primarysample==1
local wam_schoolcount_control = r(r)
quietly tab anonschoolid if treat==1 & primarysample==1
local wam_schoolcount_treat = r(r)

quietly tab anonschoolid if treat==0 & secondarysample==1
local wsem_schoolcount_control = r(r)
local idea_schoolcount_control = r(r)
quietly tab anonschoolid if treat==1 & secondarysample==1
local wsem_schoolcount_treat = r(r)
local idea_schoolcount_treat = r(r)

preserve
keep if primarysample==1
gen i=1
collapse (count) count = i, by(anonschoolid)
ameans count
local wam_nhmean : di %7.0fc r(mean_h)
restore

preserve
keep if secondarysample==1
gen i=1
collapse (count) count = i, by(anonschoolid)
ameans count
local wsem_nhmean : di %7.0fc r(mean_h)
local idea_nhmean : di %7.0fc r(mean_h)
```

```
restore

sum treat if primarysample==1
local wam_treatprop : di %7.2fc r(mean)
sum treat if secondarysample==1
local wsem_treatprop : di %7.2fc r(mean)
local idea_treatprop : di %7.2fc r(mean)

* School-level descriptives specific to FSM sub-group
preserve
keep if primarysample==1 & FSM==1
gen i=1
collapse (count) count = i, by(anonschoolid)
ameans count
local wam_fsm_nhmean : di %7.0fc r(mean_h)
restore

preserve
keep if secondarysample==1 & FSM==1
gen i=1
collapse (count) count = i, by(anonschoolid)
ameans count
local wsem_fsm_nhmean : di %7.0fc r(mean_h)
local idea_fsm_nhmean : di %7.0fc r(mean_h)
restore

sum treat if primarysample==1 & FSM==1
local wam_fsm_treatprop : di %7.2fc r(mean)
sum treat if secondarysample==1 & FSM==1
local wsem_fsm_treatprop : di %7.2fc r(mean)
local idea_fsm_treatprop : di %7.2fc r(mean)

* ICC Calculation
xtset anonschoolid
xtreg wam_score_dw
local wam_icc : di %7.2fc e(sigma_u) / (e(sigma_u) + e(sigma_e))
xtreg wsem16_score
```

```
local wsem_icc : di %7.2fc e(sigma_u) / (e(sigma_u) + e(sigma_e))
xtreg ideation_score
local idea_icc : di %7.2fc e(sigma_u) / (e(sigma_u) + e(sigma_e))

* ICC Calculation – FSM sub-group
xtset anonschoolid
xtreg wam_score_dw if FSM==1
local wam_fsm_icc : di %7.2fc e(sigma_u) / (e(sigma_u) + e(sigma_e))
xtreg wsem16_score if FSM==1
local wsem_fsm_icc : di %7.2fc e(sigma_u) / (e(sigma_u) + e(sigma_e))
xtreg ideation_score if FSM==1
local idea_fsm_icc : di %7.2fc e(sigma_u) / (e(sigma_u) + e(sigma_e))

* Group variables
cap drop FSMgroup
egen FSMgroup = mean(FSM), by(anonschoolid)
cap drop EALgroup
egen EALgroup = mean(EAL), by(anonschoolid)

* Estimate pre-post test correlations
xtset anonschoolid
xtreg wam_score_dw FSM EAL FSMgroup EALgroup
local wam_r2_pupil : di %7.2fc e(r2_w)
local wam_r2_school : di %7.2fc e(r2_b)
local wam_r2_overall : di %7.2fc e(r2_o)

xtreg wsem16_score FSM EAL FSMgroup EALgroup
local wsem_r2_pupil : di %7.2fc e(r2_w)
local wsem_r2_school : di %7.2fc e(r2_b)
local wsem_r2_overall : di %7.2fc e(r2_o)

xtreg ideation_score FSM EAL FSMgroup EALgroup
local idea_r2_pupil : di %7.2fc e(r2_w)
local idea_r2_school : di %7.2fc e(r2_b)
local idea_r2_overall : di %7.2fc e(r2_o)

* Estimate pre-post test correlations – FSM sub-group
```

```
xtset anonschoolid
xtreg wam_score_dw EAL FSMgroup EALgroup if FSM==1
local wam_fsm_r2_pupil : di %7.2fc e(r2_w)
local wam_fsm_r2_school : di %7.2fc e(r2_b)
local wam_fsm_r2_overall : di %7.2fc e(r2_o)

xtreg wsem16_score EAL FSMgroup EALgroup if FSM==1
local wsem_fsm_r2_pupil : di %7.2fc e(r2_w)
local wsem_fsm_r2_school : di %7.2fc e(r2_b)
local wsem_fsm_r2_overall : di %7.2fc e(r2_o)

xtreg ideation_score EAL FSMgroup EALgroup if FSM==1
local idea_fsm_r2_pupil : di %7.2fc e(r2_w)
local idea_fsm_r2_school : di %7.2fc e(r2_b)
local idea_fsm_r2_overall : di %7.2fc e(r2_o)

** Imbalance at baseline – as randomised
* FSM
sum FSM if treat==0 & randomisedsample==1
local fsm_controlpercent_in_rsample : di %7.1fc r(mean)*100
local fsm_controlpercent_out_rsample : di %7.1fc (1–r(mean))*100
local fsm_controlmean_rsample : di %7.2fc r(mean)
local fsm_controlsd_rsample : di %7.2fc r(sd)
local fsm_controln_rsample = r(N)
count if FSM==0 & treat==0 & randomisedsample==1
local fsm_controln_out_rsample = r(N)
count if FSM==1 & treat==0 & randomisedsample==1
local fsm_controln_in_rsample = r(N)
count if FSM>=. & treat==0 & randomisedsample==1
local fsm_controlmiss_rsample = r(N)
sum FSM if treat==1 & randomisedsample==1
local fsm_treatpercent_in_rsample : di %7.1fc r(mean)*100
local fsm_treatpercent_out_rsample : di %7.1fc (1–r(mean))*100
local fsm_treatmean_rsample : di %7.2fc r(mean)
local fsm_treatsd_rsample : di %7.2fc r(sd)
local fsm_treatn_rsample = r(N)
count if FSM==0 & treat==1 & randomisedsample==1
```

```
local fsm_treatn_out_rsample = r(N)
count if FSM==1 & treat==1 & randomisedsample==1
local fsm_treatn_in_rsample = r(N)
count if FSM>=. & treat==1 & randomisedsample==1
local fsm_treatmiss_rsample = r(N)
esizei `fsm_controln_rsample' `fsm_controlmean_rsample' `fsm_controlsd_rsample' ///
    `fsm_treatn_rsample' `fsm_treatmean_rsample' `fsm_treatsd_rsample', cohensd
local fsm_in_stddiff_rsample : di %7.3fc r(d)
local fsm_out_stddiff_rsample : di %7.3fc –r(d)

* EAL
sum EAL if treat==0 & randomisedsample==1
local eal_controlpercent_in_rsample : di %7.1fc r(mean)*100
local eal_controlpercent_out_rsample : di %7.1fc (1–r(mean))*100
local eal_controlmean_rsample : di %7.2fc r(mean)
local eal_controlsd_rsample : di %7.2fc r(sd)
local eal_controln_rsample = r(N)
count if EAL==0 & treat==0 & randomisedsample==1
local eal_controln_out_rsample = r(N)
count if EAL==1 & treat==0 & randomisedsample==1
local eal_controln_in_rsample = r(N)
count if EAL>=. & treat==0 & randomisedsample==1
local eal_controlmiss_rsample = r(N)
sum EAL if treat==1 & randomisedsample==1
local eal_treatpercent_in_rsample : di %7.1fc r(mean)*100
local eal_treatpercent_out_rsample : di %7.1fc (1–r(mean))*100
local eal_treatmean_rsample : di %7.2fc r(mean)
local eal_treatsd_rsample : di %7.2fc r(sd)
local eal_treatn_rsample = r(N)
count if EAL==0 & treat==1 & randomisedsample==1
local eal_treatn_out_rsample = r(N)
count if EAL==1 & treat==1 & randomisedsample==1
local eal_treatn_in_rsample = r(N)
count if EAL>=. & treat==1 & randomisedsample==1
local eal_treatmiss_rsample = r(N)
esizei `eal_controln_rsample' `eal_controlmean_rsample' `eal_controlsd_rsample' ///
    `eal_treatn_rsample' `eal_treatmean_rsample' `eal_treatsd_rsample', cohensd
```

```
local eal_in_stddiff_rsample : di %7.3fc r(d)
local eal_out_stddiff_rsample : di %7.3fc –r(d)

** Imbalance at baseline – as analysed
* FSM
sum FSM if treat==0 & primarysample==1
local fsm_controlpercent_in_psample : di %7.1fc r(mean)*100
local fsm_controlpercent_out_psample : di %7.1fc (1–r(mean))*100
local fsm_controlmean_psample : di %7.2fc r(mean)
local fsm_controlsd_psample : di %7.2fc r(sd)
local fsm_controln_psample = r(N)
count if FSM==0 & treat==0 & primarysample==1
local fsm_controln_out_psample = r(N)
count if FSM==1 & treat==0 & primarysample==1
local fsm_controln_in_psample = r(N)
count if FSM>=. & treat==0 & primarysample==1
local fsm_controlmiss_psample = r(N)
sum FSM if treat==1 & primarysample==1
local fsm_treatpercent_in_psample : di %7.1fc r(mean)*100
local fsm_treatpercent_out_psample : di %7.1fc (1–r(mean))*100
local fsm_treatmean_psample : di %7.2fc r(mean)
local fsm_treatsd_psample : di %7.2fc r(sd)
local fsm_treatn_psample = r(N)
count if FSM==0 & treat==1 & primarysample==1
local fsm_treatn_out_psample = r(N)
count if FSM==1 & treat==1 & primarysample==1
local fsm_treatn_in_psample = r(N)
count if FSM>=. & treat==0 & primarysample==1
local fsm_treatmiss_psample = r(N)
esizei `fsm_controln_psample' `fsm_controlmean_psample' `fsm_controlsd_psample' ///
    `fsm_treatn_psample' `fsm_treatmean_psample' `fsm_treatsd_psample', cohensd
local fsm_in_stddiff_psample : di %7.3fc r(d)
local fsm_out_stddiff_psample : di %7.3fc –r(d)

* EAL
sum EAL if treat==0 & primarysample==1
local eal_controlpercent_in_psample : di %7.1fc r(mean)*100
```

```
local eal_controlpercent_out_psample : di %7.1fc (1–r(mean))*100
local eal_controlmean_psample : di %7.2fc r(mean)
local eal_controlsd_psample : di %7.2fc r(sd)
local eal_controln_psample = r(N)
count if EAL==0 & treat==0 & primarysample==1
local eal_controln_out_psample = r(N)
count if EAL==1 & treat==0 & primarysample==1
local eal_controln_in_psample = r(N)
count if EAL>=. & treat==0 & primarysample==1
local eal_controlmiss_psample = r(N)
sum EAL if treat==1 & primarysample==1
local eal_treatpercent_in_psample : di %7.1fc r(mean)*100
local eal_treatpercent_out_psample : di %7.1fc (1–r(mean))*100
local eal_treatmean_psample : di %7.2fc r(mean)
local eal_treatsd_psample : di %7.2fc r(sd)
local eal_treatn_psample = r(N)
count if EAL==0 & treat==1 & primarysample==1
local eal_treatn_out_psample = r(N)
count if EAL==1 & treat==1 & primarysample==1
local eal_treatn_in_psample = r(N)
count if EAL>=. & treat==0 & primarysample==1
local eal_treatmiss_psample = r(N)
esizei `eal_controln_psample' `eal_controlmean_psample' `eal_controlsd_psample' ///
    `eal_treatn_psample' `eal_treatmean_psample' `eal_treatsd_psample', cohensd
local eal_in_stddiff_psample : di %7.3fc r(d)
local eal_out_stddiff_psample : di %7.3fc –r(d)

* Visualisation of outcome variables
graph twoway (kdensity wam_score_dw), scheme(plotplain) xtitle(‘Unadjusted WAM Score (ideas double weighted)’) ytitle(‘Density’)
graph export output/wam_kdensity.png, replace
graph twoway (kdensity wsem16_score), scheme(plotplain) xtitle(‘Unadjusted WSEM-16 Score’) ytitle(‘Density’)
graph export output/wsem_kdensity.png, replace
graph twoway (kdensity ideation_score), scheme(plotplain) xtitle(‘Unadjusted Ideation Score’) ytitle(‘Density’)
graph export output/ideation_kdensity.png, replace

* Descriptive statistics for outcome variables
```

```
sum wam_score_dw
local wam_mean : di %7.2fc r(mean)
local wam_sd : di %7.2fc r(sd)
local wam_n = r(N)
sum wsem16_score
local wsem_mean : di %7.2fc r(mean)
local wsem_sd : di %7.2fc r(sd)
local wsem_n = r(N)
sum ideation_score
local idea_mean : di %7.2fc r(mean)
local idea_sd : di %7.2fc r(sd)
local idea_n = r(N)

* Primary analysis
regress wam_score_dw i.treat FSM EAL FSMgroup EALgroup i.block, vce(cluster anonschoolid)
estimates store primary
local wam_treatdiff : di %7.2fc _b[1.treat]
local wam_treatdiff_upperci : di %7.2fc _b[1.treat] + 1.96*_se[1.treat]
local wam_treatdiff_lowerci : di %7.2fc _b[1.treat] – 1.96*_se[1.treat]
local wam_treatdiff_p : di %7.2fc ttail(e(df_r), abs(_b[1.treat]/_se[1.treat]))*2
margins treat, post
local wam_controlmean : di %7.2fc _b[0.treat]
local wam_controlmean_upperci : di %7.2fc _b[0.treat] + 1.96*_se[0.treat]
local wam_controlmean_lowerci : di %7.2fc _b[0.treat] – 1.96*_se[0.treat]
local wam_treatmean : di %7.2fc _b[1.treat]
local wam_treatmean_upperci : di %7.2fc _b[1.treat] + 1.96*_se[1.treat]
local wam_treatmean_lowerci : di %7.2fc _b[1.treat] – 1.96*_se[1.treat]

* Unadjusted version
regress wam_score_dw i.treat, vce(cluster anonschoolid)
local wam_treatdiff_unadj : di %7.2fc _b[1.treat]
local wam_treatdiff_unadj_upperci : di %7.2fc _b[1.treat] + 1.96*_se[1.treat]
local wam_treatdiff_unadj_lowerci : di %7.2fc _b[1.treat] – 1.96*_se[1.treat]
local wam_treatdiff_unadj_p : di %7.2fc ttail(e(df_r), abs(_b[1.treat]/_se[1.treat]))*2
margins treat, post
local wam_controlmean_unadj : di %7.2fc _b[0.treat]
local wam_controlmean_unadj_upperci : di %7.2fc _b[0.treat] + 1.96*_se[0.treat]
```

```
local wam_controlmean_unadj_lowerci : di %7.2fc _b[0.treat] – 1.96*_se[0.treat]
local wam_treatmean_unadj : di %7.2fc _b[1.treat]
local wam_treatmean_unadj_upperci : di %7.2fc _b[1.treat] + 1.96*_se[1.treat]
local wam_treatmean_unadj_lowerci : di %7.2fc _b[1.treat] – 1.96*_se[1.treat]

* Graphical version of primary analysis
regress wam_score_dw FSM EAL FSMgroup EALgroup i.block, vce(cluster anonschoolid)
cap drop wam_analysis_resid
predict wam_analysis_resid, resid
graph twoway ///
    (kdensity wam_score_dw if treat==1) ///
    (kdensity wam_score_dw if treat==0) ///
    , scheme(plotplain) legend(order(1 'Treatment' 2 'Control')) xtitle('Unadjusted WAM Score') ytitle('Density')
graph export output/wam_kdensity_treat_uncond.png, replace
graph twoway ///
    (kdensity wam_analysis_resid if treat==1) ///
    (kdensity wam_analysis_resid if treat==0) ///
    , scheme(plotplain) legend(order(1 'Treatment' 2 'Control')) xtitle('Adjusted WAM Score') ytitle('Density')
graph export output/wam_kdensity_treat_cond.png, replace

* Secondary analysis – WSEM
regress wsem16_score i.treat FSM EAL FSMgroup EALgroup i.block, vce(cluster anonschoolid)
local wsem_treatdiff : di %7.2fc _b[1.treat]
local wsem_treatdiff_upperci : di %7.2fc _b[1.treat] + 1.96*_se[1.treat]
local wsem_treatdiff_lowerci : di %7.2fc _b[1.treat] – 1.96*_se[1.treat]
local wsem_treatdiff_p : di %7.2fc ttail(e(df_r), abs(_b[1.treat]/_se[1.treat]))*2
margins treat, post
local wsem_controlmean : di %7.2fc _b[0.treat]
local wsem_controlmean_upperci : di %7.2fc _b[0.treat] + 1.96*_se[0.treat]
local wsem_controlmean_lowerci : di %7.2fc _b[0.treat] – 1.96*_se[0.treat]
local wsem_treatmean : di %7.2fc _b[1.treat]
local wsem_treatmean_upperci : di %7.2fc _b[1.treat] + 1.96*_se[1.treat]
local wsem_treatmean_lowerci : di %7.2fc _b[1.treat] – 1.96*_se[1.treat]

* Unadjusted version
regress wsem16_score i.treat, vce(cluster anonschoolid)
local wsem_treatdiff_unadj : di %7.2fc _b[1.treat]
```

```
local wsem_treatdiff_unadj_upperci : di %7.2fc _b[1.treat] + 1.96*_se[1.treat]
local wsem_treatdiff_unadj_lowerci : di %7.2fc _b[1.treat] – 1.96*_se[1.treat]
local wsem_treatdiff_unadj_p : di %7.2fc ttail(e(df_r), abs(_b[1.treat]/_se[1.treat]))*2
margins treat, post
local wsem_controlmean_unadj : di %7.2fc _b[0.treat]
local wsem_controlmean_unadj_upperci : di %7.2fc _b[0.treat] + 1.96*_se[0.treat]
local wsem_controlmean_unadj_lowerci : di %7.2fc _b[0.treat] – 1.96*_se[0.treat]
local wsem_treatmean_unadj : di %7.2fc _b[1.treat]
local wsem_treatmean_unadj_upperci : di %7.2fc _b[1.treat] + 1.96*_se[1.treat]
local wsem_treatmean_unadj_lowerci : di %7.2fc _b[1.treat] – 1.96*_se[1.treat]

* Graphical version of secondary analysis – WSEM
regress wsem16_score FSM EAL FSMgroup EALgroup i.block, vce(cluster anonschoolid)
cap drop wsem_analysis_resid
predict wsem_analysis_resid, resid
graph twoway ///
    (kdensity wsem16_score if treat==1) ///
    (kdensity wsem16_score if treat==0) ///
    , scheme(plotplain) legend(order(1 'Treatment' 2 'Control')) xtitle('Unadjusted WSEM Score') ytitle('Density')
graph export output/wsem_kdensity_treat_uncond.png, replace
graph twoway ///
    (kdensity wsem_analysis_resid if treat==1) ///
    (kdensity wsem_analysis_resid if treat==0) ///
    , scheme(plotplain) legend(order(1 'Treatment' 2 'Control')) xtitle('Adjusted WSEM Score') ytitle('Density')
graph export output/wsem_kdensity_treat_cond.png, replace

* Secondary analysis – Ideation
regress ideation_score i.treat FSM EAL FSMgroup EALgroup i.block, vce(cluster anonschoolid)
local idea_treatdiff : di %7.2fc _b[1.treat]
local idea_treatdiff_upperci : di %7.2fc _b[1.treat] + 1.96*_se[1.treat]
local idea_treatdiff_lowerci : di %7.2fc _b[1.treat] – 1.96*_se[1.treat]
local idea_treatdiff_p : di %7.2fc ttail(e(df_r), abs(_b[1.treat]/_se[1.treat]))*2
margins treat, post
local idea_controlmean : di %7.2fc _b[0.treat]
local idea_controlmean_upperci : di %7.2fc _b[0.treat] + 1.96*_se[0.treat]
local idea_controlmean_lowerci : di %7.2fc _b[0.treat] – 1.96*_se[0.treat]
local idea_treatmean : di %7.2fc _b[1.treat]
```

```
local idea_treatmean_upperci : di %7.2fc _b[1.treat] + 1.96*_se[1.treat]
local idea_treatmean_lowerci : di %7.2fc _b[1.treat] – 1.96*_se[1.treat]

regress ideation_score i.treat, vce(cluster anonschoolid)
local idea_treatdiff_unadj : di %7.2fc _b[1.treat]
local idea_treatdiff_unadj_upperci : di %7.2fc _b[1.treat] + 1.96*_se[1.treat]
local idea_treatdiff_unadj_lowerci : di %7.2fc _b[1.treat] – 1.96*_se[1.treat]
local idea_treatdiff_unadj_p : di %7.2fc ttail(e(df_r), abs(_b[1.treat]/_se[1.treat]))*2
margins treat, post
local idea_controlmean_unadj : di %7.2fc _b[0.treat]
local idea_controlmean_unadj_upperci : di %7.2fc _b[0.treat] + 1.96*_se[0.treat]
local idea_controlmean_unadj_lowerci : di %7.2fc _b[0.treat] – 1.96*_se[0.treat]
local idea_treatmean_unadj : di %7.2fc _b[1.treat]
local idea_treatmean_unadj_upperci : di %7.2fc _b[1.treat] + 1.96*_se[1.treat]
local idea_treatmean_unadj_lowerci : di %7.2fc _b[1.treat] – 1.96*_se[1.treat]

* Graphical version of secondary analysis – ideation
regress ideation_score FSM EAL FSMgroup EALgroup i.block, vce(cluster anonschoolid)
cap drop ideation_analysis_resid
predict ideation_analysis_resid, resid
graph twoway ///
    (kdensity ideation_score if treat==1) ///
    (kdensity ideation_score if treat==0) ///
    , scheme(plotplain) legend(order(1 'Treatment' 2 'Control')) xtitle('Unadjusted Ideation Score') ytitle('Density')
graph export output/ideation_kdensity_treat_uncond.png, replace
graph twoway ///
    (kdensity ideation_analysis_resid if treat==1) ///
    (kdensity ideation_analysis_resid if treat==0) ///
    , scheme(plotplain) legend(order(1 'Treatment' 2 'Control')) xtitle('Adjusted Ideation Score') ytitle('Density')
graph export output/ideation_kdensity_treat_cond.png, replace

* Effect size calculations – primary analysis
count if primarysample==1 & treat==0
local wam_controln = r(N)
count if primarysample==1 & treat==1
local wam_treatn = r(N)
quietly summarize wam_score_dw if primarysample==1 & treat==0
```

```
local wam_controlsd : di %7.2fc r(sd)
quietly summarize wam_score_dw if primarysample==1 & treat==1
local wam_treatsd : di %7.2fc r(sd)
local wam_sdpooled = sqrt(((`wam_controln'–1)*(`wam_controlsd'^2) + (`wam_treatn'–1)*(`wam_treatsd'^2))/(`wam_controln'+`wam_treatn'–2))
local wam_sdpooled : di %7.2fc `wam_sdpooled'

local wam_j = (exp(lngamma((`wam_treatn'+`wam_controln'+2)/2))) / (((`wam_treatn'+`wam_controln'+2)/2)^.5 * exp(lngamma((`wam_treatn'+`wam_controln'+2–1)/2)))
if `wam_j' == . {
   local wam_j = 1–(3/((4*(`wam_treatn'+`wam_controln'))–9))
   }
local wam_g : di %7.2fc `wam_j'*(`wam_treatdiff'/`wam_sdpooled')
local wam_g_upperci : di %7.2fc `wam_j'*(`wam_treatdiff_upperci'/`wam_sdpooled')
local wam_g_lowerci : di %7.2fc `wam_j'*(`wam_treatdiff_lowerci'/`wam_sdpooled')
local wam_j : di %7.5fc `wam_j'

* Effect size calculations – secondary analysis – WSEM
count if wsem16_score<. & secondarysample==1 & treat==0
local wsem_controln = r(N)
count if wsem16_score<. & secondarysample==1 & treat==1
local wsem_treatn = r(N)
quietly summarize wsem16_score if secondarysample==1 & treat==0
local wsem_controlsd : di %7.2fc r(sd)
quietly summarize wsem16_score if secondarysample==1 & treat==1
local wsem_treatsd : di %7.2fc r(sd)
local wsem_sdpooled = sqrt(((`wsem_controln'–1)*(`wsem_controlsd'^2) + (`wsem_treatn'–1)*(`wsem_treatsd'^2))/(`wsem_controln'+`wsem_treatn'–2))
local wsem_sdpooled : di %7.2fc `wsem_sdpooled'

local wsem_j = (exp(lngamma((`wsem_treatn'+`wsem_controln'+2)/2))) / (((`wsem_treatn'+`wsem_controln'+2)/2)^.5 * exp(lngamma((`wsem_treatn'+`wsem_controln'+2–1)/2)))
if `wsem_j' == . {
   local wsem_j = 1–(3/((4*(`wsem_treatn'+`wsem_controln'))–9))
   }
local wsem_g : di %7.2fc `wsem_j'*(`wsem_treatdiff'/`wsem_sdpooled')
local wsem_g_upperci : di %7.2fc `wsem_j'*(`wsem_treatdiff_upperci'/`wsem_sdpooled')
local wsem_g_lowerci : di %7.2fc `wsem_j'*(`wsem_treatdiff_lowerci'/`wsem_sdpooled')
```

```
local wsem_j : di %7.5fc `wsem_j'

* Effect size calculations – secondary analysis – Ideation
count if ideation_score<. & secondarysample==1 & treat==0
local idea_controln = r(N)
count if ideation_score<. & secondarysample==1 & treat==1
local idea_treatn = r(N)
quietly summarize ideation_score if secondarysample==1 & treat==0
local idea_controlsd : di %7.2fc r(sd)
quietly summarize ideation_score if secondarysample==1 & treat==1
local idea_treatsd : di %7.2fc r(sd)
local idea_sdpooled = sqrt(((`idea_controln'–1)*(`idea_controlsd'^2) + (`idea_treatn'–
1)*(`idea_treatsd'^2))/(`idea_controln'+`idea_treatn'–2))
local idea_sdpooled : di %7.2fc `idea_sdpooled'

local idea_j = (exp(lngamma((`idea_treatn'+`idea_controln'+2)/2))) / (((`idea_treatn'+`idea_controln'+2)/2)^.5 *
exp(lngamma((`idea_treatn'+`idea_controln'+2–1)/2)))
if `idea_j' == . {
   local idea_j = 1–(3/((4*(`idea_treatn'+`idea_controln'))–9))
   }
local idea_g : di %7.2fc `idea_j'*(`idea_treatdiff'/`idea_sdpooled')
local idea_g_upperci : di %7.2fc `idea_j'*(`idea_treatdiff_upperci'/`idea_sdpooled')
local idea_g_lowerci : di %7.2fc `idea_j'*(`idea_treatdiff_lowerci'/`idea_sdpooled')
local idea_j : di %7.5fc `idea_j'

* Complier analysis
corr treat comply
local complier_r : di %7.2fc r(rho)
regress comply FSM EAL FSMgroup EALgroup i.block treat if primarysample==1, vce(cluster anonschoolid)
local complier_F : di %7.0f e(F)
local complier_df_m = e(df_m)
local complier_df_r = e(df_r)
ivregress 2sls wam_score_dw FSM EAL FSMgroup EALgroup i.block (i.comply = treat), vce(cluster anonschoolid)
local complier_treatdiff : di %7.2fc _b[1.comply]
local complier_treatdiff_upperci : di %7.2fc _b[1.comply] + 1.96*_se[1.comply]
local complier_treatdiff_lowerci : di %7.2fc _b[1.comply] – 1.96*_se[1.comply]
local complier_treatdiff_p : di %7.2fc normal(–abs(_b[1.comply]/_se[1.comply]))*2
margins comply, post
```

```
local complier_controlmean : di %7.2fc _b[0.comply]
local complier_controlmean_upperci : di %7.2fc _b[0.comply] + 1.96*_se[0.comply]
local complier_controlmean_lowerci : di %7.2fc _b[0.comply] – 1.96*_se[0.comply]
local complier_treatmean : di %7.2fc _b[1.comply]
local complier_treatmean_upperci : di %7.2fc _b[1.comply] + 1.96*_se[1.comply]
local complier_treatmean_lowerci : di %7.2fc _b[1.comply] – 1.96*_se[1.comply]

* Effect size calculations – complier analysis
count if primarysample==1 & comply<. & treat==0
local complier_controln = r(N)
count if primarysample==1 & comply<. & treat==1
local complier_treatn = r(N)
count if primarysample==1 & comply<.
local complier_n = r(N)
quietly summarize wam_score_dw if primarysample==1 & comply<. & treat==0
local complier_controlsd : di %7.2fc r(sd)
quietly summarize wam_score_dw if primarysample==1 & comply<. & treat==1
local complier_treatsd : di %7.2fc r(sd)
local complier_sdpooled = sqrt(((`complier_controln'–1)*(`complier_controlsd'^2) + (`complier_treatn'–
1)*(`complier_treatsd'^2))/(`complier_controln'+`complier_treatn'–2))
local complier_sdpooled : di %7.2fc `complier_sdpooled'

local complier_j = (exp(lngamma((`complier_treatn'+`complier_controln'+2)/2))) /
(((`complier_treatn'+`complier_controln'+2)/2)^.5 * exp(lngamma((`complier_treatn'+`complier_controln'+2–1)/2)))
if `complier_j' == . {
    local complier_j = 1–(3/((4*(`complier_treatn'+`complier_controln'))–9))
   }
local complier_g : di %7.2fc `complier_j'*(`complier_treatdiff'/`complier_sdpooled')
local complier_g_upperci : di %7.2fc `complier_j'*(`complier_treatdiff_upperci'/`complier_sdpooled')
local complier_g_lowerci : di %7.2fc `complier_j'*(`complier_treatdiff_lowerci'/`complier_sdpooled')

* FSM check for interaction then run for EVERFSM_6_P sub-group – primary analysis
regress wam_score_dw i.treat i.FSM treat#FSM EAL FSMgroup EALgroup i.block, vce(cluster anonschoolid)
local fsmwam_interaction_p : di %7.2fc ttail(e(df_r), abs(_b[1.treat#1.FSM]/_se[1.treat#1.FSM]))*2

regress wam_score_dw i.treat EAL FSMgroup EALgroup i.block if FSM==1, vce(cluster anonschoolid)

local fsmwam_treatdiff : di %7.2fc _b[1.treat]
```

```
local fsmwam_treatdiff_upperci : di %7.2fc _b[1.treat] + 1.96*_se[1.treat]
local fsmwam_treatdiff_lowerci : di %7.2fc _b[1.treat] – 1.96*_se[1.treat]
local fsmwam_treatdiff_p : di %7.2fc ttail(e(df_r), abs(_b[1.treat]/_se[1.treat]))*2
margins treat, post
local fsmwam_controlmean : di %7.2fc _b[0.treat]
local fsmwam_controlmean_upperci : di %7.2fc _b[0.treat] + 1.96*_se[0.treat]
local fsmwam_controlmean_lowerci : di %7.2fc _b[0.treat] – 1.96*_se[0.treat]
local fsmwam_treatmean : di %7.2fc _b[1.treat]
local fsmwam_treatmean_upperci : di %7.2fc _b[1.treat] + 1.96*_se[1.treat]
local fsmwam_treatmean_lowerci : di %7.2fc _b[1.treat] – 1.96*_se[1.treat]

* FSM check for interaction then run for EVERFSM_6_P sub-group – secondary analysis (WSEM)
regress wsem16_score i.treat i.FSM treat#FSM EAL FSMgroup EALgroup i.block, vce(cluster anonschoolid)
local fsmwsem_interaction_p : di %7.2fc ttail(e(df_r), abs(_b[1.treat#1.FSM]/_se[1.treat#1.FSM]))*2

regress wam_score_dw i.treat EAL FSMgroup EALgroup i.block if FSM==1, vce(cluster anonschoolid)

local fsmwsem_treatdiff : di %7.2fc _b[1.treat]
local fsmwsem_treatdiff_upperci : di %7.2fc _b[1.treat] + 1.96*_se[1.treat]
local fsmwsem_treatdiff_lowerci : di %7.2fc _b[1.treat] – 1.96*_se[1.treat]
local fsmwsem_treatdiff_p : di %7.2fc ttail(e(df_r), abs(_b[1.treat]/_se[1.treat]))*2
margins treat, post
local fsmwsem_controlmean : di %7.2fc _b[0.treat]
local fsmwsem_controlmean_upperci : di %7.2fc _b[0.treat] + 1.96*_se[0.treat]
local fsmwsem_controlmean_lowerci : di %7.2fc _b[0.treat] – 1.96*_se[0.treat]
local fsmwsem_treatmean : di %7.2fc _b[1.treat]
local fsmwsem_treatmean_upperci : di %7.2fc _b[1.treat] + 1.96*_se[1.treat]
local fsmwsem_treatmean_lowerci : di %7.2fc _b[1.treat] – 1.96*_se[1.treat]

* FSM check for interaction then run for EVERFSM_6_P sub-group – secondary analysis (Ideation)
regress ideation_score i.treat i.FSM treat#FSM EAL FSMgroup EALgroup i.block, vce(cluster anonschoolid)
local fsmidea_interaction_p : di %7.2fc ttail(e(df_r), abs(_b[1.treat#1.FSM]/_se[1.treat#1.FSM]))*2

regress ideation_score i.treat EAL FSMgroup EALgroup i.block if FSM==1, vce(cluster anonschoolid)

local fsmidea_treatdiff : di %7.2fc _b[1.treat]
local fsmidea_treatdiff_upperci : di %7.2fc _b[1.treat] + 1.96*_se[1.treat]
```

```
local fsmidea_treatdiff_lowerci : di %7.2fc _b[1.treat] – 1.96*_se[1.treat]
local fsmidea_treatdiff_p : di %7.2fc ttail(e(df_r), abs(_b[1.treat]/_se[1.treat]))*2
margins treat, post
local fsmidea_controlmean : di %7.2fc _b[0.treat]
local fsmidea_controlmean_upperci : di %7.2fc _b[0.treat] + 1.96*_se[0.treat]
local fsmidea_controlmean_lowerci : di %7.2fc _b[0.treat] – 1.96*_se[0.treat]
local fsmidea_treatmean : di %7.2fc _b[1.treat]
local fsmidea_treatmean_upperci : di %7.2fc _b[1.treat] + 1.96*_se[1.treat]
local fsmidea_treatmean_lowerci : di %7.2fc _b[1.treat] – 1.96*_se[1.treat]

* Effect size calculations – FSM sub-group analysis
count if primarysample==1 & FSM==1 & treat==0
local primarysample_fsm_control_n = r(N)
count if primarysample==1 & FSM==1 & treat==1
local primarysample_fsm_treat_n = r(N)
count if primarysample==1 & FSM==1
local primarysample_fsm_n = r(N)

count if secondarysample==1 & FSM==1 & treat==0
local secondarysample_fsm_control_n = r(N)
count if secondarysample==1 & FSM==1 & treat==1
local secondarysample_fsm_treat_n = r(N)
count if secondarysample==1 & FSM==1
local secondarysample_fsm_n = r(N)

* Primary
quietly summarize wam_score_dw if primarysample==1 & FSM==1 & treat==0
local fsmwam_controlsd : di %7.2fc r(sd)
quietly summarize wam_score_dw if primarysample==1 & FSM==1 & treat==1
local fsmwam_treatsd : di %7.2fc r(sd)
local fsmwam_sdpooled = sqrt(((`primarysample_fsm_control_n'–1)*(`fsmwam_controlsd'^2) +
(`primarysample_fsm_treat_n'–
1)*(`fsmwam_treatsd'^2))/(`primarysample_fsm_control_n'+`primarysample_fsm_treat_n'–2))
local fsmwam_sdpooled : di %7.2fc `fsmwam_sdpooled'

local fsmwam_j = (exp(lngamma((`primarysample_fsm_treat_n'+`primarysample_fsm_control_n'+2)/2))) /
(((`primarysample_fsm_treat_n'+`primarysample_fsm_control_n'+2)/2)^.5 *
exp(lngamma((`primarysample_fsm_treat_n'+`primarysample_fsm_control_n'+2–1)/2)))
```

```
if `fsmwam_j' == . {
    local fsmwam_j = 1–(3/((4*(`primarysample_fsm_treat_n'+`primarysample_fsm_control_n'))–9))
    }
local fsmwam_g : di %7.2fc `fsmwam_j'*(`fsmwam_treatdiff'/`fsmwam_sdpooled')
local fsmwam_g_upperci : di %7.2fc `fsmwam_j'*(`fsmwam_treatdiff_upperci'/`fsmwam_sdpooled')
local fsmwam_g_lowerci : di %7.2fc `fsmwam_j'*(`fsmwam_treatdiff_lowerci'/`fsmwam_sdpooled')

* Secondary (WSEM)
quietly summarize wsem16_score if secondarysample==1 & FSM==1 & treat==0
local fsmwsem_controlsd : di %7.2fc r(sd)
quietly summarize wsem16_score if secondarysample==1 & FSM==1 & treat==1
local fsmwsem_treatsd : di %7.2fc r(sd)
local fsmwsem_sdpooled = sqrt(((`secondarysample_fsm_control_n'–1)*(`fsmwsem_controlsd'^2) +
(`secondarysample_fsm_treat_n'–
1)*(`fsmwsem_treatsd'^2))/(`secondarysample_fsm_control_n'+`secondarysample_fsm_treat_n'–2))
local fsmwsem_sdpooled : di %7.2fc `fsmwsem_sdpooled'

local fsmwsem_j = (exp(lngamma((`secondarysample_fsm_treat_n'+`secondarysample_fsm_control_n'+2)/2))) /
(((`secondarysample_fsm_treat_n'+`secondarysample_fsm_control_n'+2)/2)^.5 *
exp(lngamma((`secondarysample_fsm_treat_n'+`secondarysample_fsm_control_n'+2–1)/2)))
if `fsmwsem_j' == . {
    local fsmwsem_j = 1–(3/((4*(`secondarysample_fsm_treat_n'+`secondarysample_fsm_control_n'))–9))
    }
local fsmwsem_g : di %7.2fc `fsmwsem_j'*(`fsmwsem_treatdiff'/`fsmwsem_sdpooled')
local fsmwsem_g_upperci : di %7.2fc `fsmwsem_j'*(`fsmwsem_treatdiff_upperci'/`fsmwsem_sdpooled')
local fsmwsem_g_lowerci : di %7.2fc `fsmwsem_j'*(`fsmwsem_treatdiff_lowerci'/`fsmwsem_sdpooled')

* Secondary (Ideation)
quietly summarize ideation_score if secondarysample==1 & FSM==1 & treat==0
local fsmidea_controlsd : di %7.2fc r(sd)
quietly summarize ideation_score if secondarysample==1 & FSM==1 & treat==1
local fsmidea_treatsd : di %7.2fc r(sd)
local fsmidea_sdpooled = sqrt(((`secondarysample_fsm_control_n'–1)*(`fsmidea_controlsd'^2) +
(`secondarysample_fsm_treat_n'–
1)*(`fsmidea_treatsd'^2))/(`secondarysample_fsm_control_n'+`secondarysample_fsm_treat_n'–2))
local fsmidea_sdpooled : di %7.2fc `fsmidea_sdpooled'
```

```
local fsmidea_j = (exp(lngamma((`secondarysample_fsm_treat_n'+`secondarysample_fsm_control_n'+2)/2))) /
(((`secondarysample_fsm_treat_n'+`secondarysample_fsm_control_n'+2)/2)^.5 *
exp(lngamma((`secondarysample_fsm_treat_n'+`secondarysample_fsm_control_n'+2–1)/2)))
if `fsmidea_j' == . {
   local fsmidea_j = 1–(3/((4*(`secondarysample_fsm_treat_n'+`secondarysample_fsm_control_n'))–9))
   }
local fsmidea_g : di %7.2fc `fsmidea_j'*(`fsmidea_treatdiff'/`fsmidea_sdpooled')
local fsmidea_g_upperci : di %7.2fc `fsmidea_j'*(`fsmidea_treatdiff_upperci'/`fsmidea_sdpooled')
local fsmidea_g_lowerci : di %7.2fc `fsmidea_j'*(`fsmidea_treatdiff_lowerci'/`fsmidea_sdpooled')

* Missing data analysis – create missing indicator variables
cap drop wam_score_dw_miss
gen wam_score_dw_miss = 0
replace wam_score_dw_miss = 1 if wam_score_dw>=.
logit wam_score_dw_miss i.treat FSM FSMgroup EAL EALgroup i.block, vce(cluster anonschoolid)
cap drop wsem16_score_miss
gen wsem16_score_miss = 0
replace wsem16_score_miss = 1 if wsem16_score>=.
logit wam_score_dw_miss i.treat FSM FSMgroup EAL EALgroup i.block, vce(cluster anonschoolid)

* Missing data analysis – only outcome variable missing
regress wam_score_dw i.treat FSM FSMgroup EAL EALgroup i.block, vce(cluster anonschoolid)
* [IF THERE WERE ANY OTHER STATISTICALLY SIGNIFICANT PREDICTORS OF wam_score_dw_miss NOT
ALREADY HERE WE'D ADD THEM AND REPORT THIS AS A ROBUSTNESS CHECK MODEL]

* Missing data analysis – predictor variables
* [NO MISSING PREDICTOR VARIABLE DATA IN ANALYSIS SAMPLE]

* CREATE TABLES

* Balance table – Randomisation sample
cap file close baltab
file open baltab using output/balance_rsample.csv, write replace
file write baltab ‘    , , Control, , Intervention’ _n
file write baltab ‘Pupil-level (categorical), National-level mean, n/N (missing), Count (%), n/N (missing), Count (%),
Standardised difference’ _n
file write baltab ‘Ever FSM, ${natmean_fsm}, `fsm_controln_rsample' / `asrandom_schoolcount_control'
(`fsm_controlmiss_rsample'), `fsm_controln_in_rsample' (`fsm_controlpercent_in_rsample'), ‘
```

file write baltab ‘`fsm_treatn_rsample' / `asrandom_schoolcount_treat' (`fsm_treatmiss_rsample'), `fsm_treatn_in_rsample' (`fsm_treatpercent_in_rsample'), `fsm_in_stddiff_rsample'‘ _n

file write baltab ‘Non-Ever FSM, ${natmean_nonfsm}, `fsm_controln_rsample' / `asrandom_schoolcount_control' (`fsm_controlmiss_rsample'), `fsm_controln_out_rsample' (`fsm_controlpercent_out_rsample'), ‘

file write baltab ‘`fsm_treatn_rsample' / `asrandom_schoolcount_treat' (`fsm_treatmiss_rsample'), `fsm_treatn_out_rsample' (`fsm_treatpercent_out_rsample'), `fsm_out_stddiff_rsample'‘ _n

file write baltab ‘EAL, ${natmean_eal}, `eal_controln_rsample' / `asrandom_schoolcount_control' (`eal_controlmiss_rsample'), `eal_controln_in_rsample' (`eal_controlpercent_in_rsample'), ‘

file write baltab ‘`eal_treatn_rsample' / `asrandom_schoolcount_treat' (`eal_treatmiss_rsample'), `eal_treatn_in_rsample' (`eal_treatpercent_in_rsample'), `eal_in_stddiff_rsample'‘ _n

file write baltab ‘Non-EAL, ${natmean_noneal}, `eal_controln_rsample' / `asrandom_schoolcount_control' (`eal_controlmiss_rsample'), `eal_controln_out_rsample' (`eal_controlpercent_out_rsample'), ‘

file write baltab ‘`eal_treatn_rsample' / `asrandom_schoolcount_treat' (`eal_treatmiss_rsample'), `eal_treatn_out_rsample' (`eal_treatpercent_out_rsample'), `eal_out_stddiff_rsample'‘ _n

file close baltab

* Balance table - Primary analysis sample

cap file close baltab

file open baltab using output/balance_psample.csv, write replace

file write baltab ‘ , , Control, , Intervention’ _n

file write baltab ‘Pupil-level (categorical), National-level mean, n/N (missing), Count (%), n/N (missing), Count (%), Standardised difference’ _n

file write baltab ‘Ever FSM, ${natmean_fsm}, `fsm_controln_psample' / `asrandom_schoolcount_control' (`fsm_controlmiss_psample'), `fsm_controln_in_psample' (`fsm_controlpercent_in_psample'), ‘

file write baltab ‘`fsm_treatn_psample' / `asrandom_schoolcount_treat' (`fsm_treatmiss_psample'), `fsm_treatn_in_psample' (`fsm_treatpercent_in_psample'), `fsm_in_stddiff_psample'‘ _n

file write baltab ‘Non-Ever FSM, ${natmean_nonfsm}, `fsm_controln_psample' / `asrandom_schoolcount_control' (`fsm_controlmiss_psample'), `fsm_controln_out_psample' (`fsm_controlpercent_out_psample'), ‘

file write baltab ‘`fsm_treatn_psample' / `asrandom_schoolcount_treat' (`fsm_treatmiss_psample'), `fsm_treatn_out_psample' (`fsm_treatpercent_out_psample'), `fsm_out_stddiff_psample'‘ _n

file write baltab ‘EAL, ${natmean_eal}, `eal_controln_psample' / `asrandom_schoolcount_control' (`eal_controlmiss_psample'), `eal_controln_in_psample' (`eal_controlpercent_in_psample'), ‘

file write baltab ‘`eal_treatn_psample' / `asrandom_schoolcount_treat' (`eal_treatmiss_psample'), `eal_treatn_in_psample' (`eal_treatpercent_in_psample'), `eal_in_stddiff_psample'‘ _n

file write baltab ‘Non-EAL, ${natmean_noneal}, `eal_controln_psample' / `asrandom_schoolcount_control' (`eal_controlmiss_psample'), `eal_controln_out_psample' (`eal_controlpercent_out_psample'), ‘

file write baltab ‘`eal_treatn_psample' / `asrandom_schoolcount_treat' (`eal_treatmiss_psample'), `eal_treatn_out_psample' (`eal_treatpercent_out_psample'), `eal_out_stddiff_psample'‘ _n

file close baltab

* Outcome measure descriptive statistics

cap file close outcomes

```
file open outcomes using output/outcomes.csv, write replace
file write outcomes “Outcome, Mean, SD, ICC, N” _n
file write outcomes “WAM Score (ideas scale double weighted), `wam_mean', `wam_sd', `wam_icc', `wam_n'‘ _n
file write outcomes “WSEM-16 Score, `wsem_mean', `wsem_sd', `wsem_icc', `wsem_n'‘ _n
file write outcomes “Ideation Score, `idea_mean', `idea_sd', `idea_icc', `idea_n'‘ _n
file close outcomes

* Primary analysis
cap file close primary
file open primary using output/primary.csv, write replace
file write primary “Unadjusted means, Full sample , Control group, , Intervention group, , Effect size calculation, , ‘ _n
file write primary “Outcome, n (missing), Mean (95% CI), n (missing), Mean (95% CI), n (missing), Total n (intervention; control), Hedges' g (95% CI), p-value” _n
file write primary “WAM Score (ideas scale double weighted), `primarymeasure_n' (`primarymeasure_miss'), ‘
file write primary ‘`wam_controlmean_unadj' (`wam_controlmean_unadj_lowerci'; `wam_controlmean_unadj_upperci'), `primarymeasure_control_n' (`primarymeasure_control_miss'), ‘
file write primary ‘`wam_treatmean_unadj' (`wam_treatmean_unadj_lowerci'; `wam_treatmean_unadj_upperci'), `primarymeasure_treat_n' (`primarymeasure_treat_miss'), ‘
file write primary ‘`primarysample_n' (`primarysample_treat_n'; `primarysample_control_n'), `wam_g' (`wam_g_lowerci'; `wam_g_upperci'), `wam_treatdiff_p'‘ _n
file close primary

* Secondary analysis
cap file close secondary
file open secondary using output/secondary.csv, write replace
file write secondary “Unadjusted means, Full sample , Control group, , Intervention group, , Effect size, , ‘ _n
file write secondary “Outcome, n (missing), Mean (95% CI), n (missing), Mean (95% CI), n (missing), Total n (intervention; control), Hedges' g (95% CI), p-value” _n
file write secondary “WSEM Score, `wsem_n' (`wsem_miss'), ‘
file write secondary ‘`wsem_controlmean_unadj' (`wsem_controlmean_unadj_lowerci'; `wsem_controlmean_unadj_upperci'), `wsem_control_n' (`wsem_control_miss'), ‘
file write secondary ‘`wsem_treatmean_unadj' (`wsem_treatmean_unadj_lowerci'; `wsem_treatmean_unadj_upperci'), `wsem_treat_n' (`wsem_treat_miss'), ‘
file write secondary ‘`secondarysample_n' (`secondarysample_treat_n'; `secondarysample_control_n'), `wsem_g' (`wsem_g_lowerci'; `wsem_g_upperci'), `wsem_treatdiff_p'‘ _n
file write secondary “Ideation Score, `idea_n' (`idea_miss'), ‘
file write secondary ‘`idea_controlmean_unadj' (`idea_controlmean_unadj_lowerci'; `idea_controlmean_unadj_upperci'), `idea_control_n' (`idea_control_miss'), ‘
file write secondary ‘`idea_treatmean_unadj' (`idea_treatmean_unadj_lowerci'; `idea_treatmean_unadj_upperci'), `idea_treat_n' (`idea_treat_miss'), ‘
```

file write secondary ‘`secondarysample_n' (`secondarysample_treat_n'; `secondarysample_control_n'), `idea_g' (`idea_g_lowerci'; `idea_g_upperci'), `idea_treatdiff_p'‘ _n

file close secondary

* Sub-group analyses

cap file close sub-group

file open sub-group using output/sub-group.csv, write replace

file write sub-group ‘Model, Hedges' g (95% CI), N, p-value of interaction term, p-value of treatment variable ‘ _n

file write sub-group ‘WAM FSM sub-group analysis, `fsmwam_g' (`fsmwam_g_lowerci'; `fsmwam_g_upperci'), `primarysample_fsm_n', `fsmwam_interaction_p', `fsmwam_treatdiff_p'‘ _n

file write sub-group ‘WSEM FSM sub-group analysis, `fsmwsem_g' (`fsmwsem_g_lowerci'; `fsmwsem_g_upperci'), `secondarysample_fsm_n', `fsmwsem_interaction_p', `fsmwsem_treatdiff_p'‘ _n

file write sub-group ‘Ideation FSM sub-group analysis, `fsmidea_g' (`fsmidea_g_lowerci'; `fsmidea_g_upperci'), `secondarysample_fsm_n', `fsmidea_interaction_p', `fsmidea_treatdiff_p'‘ _n

file close sub-group

* Compliance analysis

cap file close sub-group

file open sub-group using output/compliance.csv, write replace

file write sub-group ‘Model, Hedges' g (95% CI), N, First stage F test, Compliance/treatment correlation, p-value of treatment variable ‘ _n

file write sub-group `’Compliance analysis, `complier_g' ‘(`complier_g_lowerci', `complier_g_upperci')’, `complier_n', ‘F(`complier_df_m', `complier_df_r') = `complier_F'‘, `complier_r', `complier_treatdiff_p'‘' _n

file close sub-group

* Attrition table

cap file close attrition

file open attrition using output/attrition.csv, write replace

file write attrition ‘ , , Intervention, Control, Total’ _n

file write attrition ‘Number of pupils, Randomised, `randomisedsample_treat_n', `randomisedsample_control_n', `randomisedsample_n'‘ _n

file write attrition ‘ , Analysed, `primarysample_treat_n', `primarysample_control_n', `primarysample_n'‘ _n

file write attrition ‘Pupil attrition, Number, `primaryattrition_treat_n', `primaryattrition_control_n', `primaryattrition_n'‘ _n

file write attrition ‘(from randomisation to analysis) , Percentage, `primaryattrition_treat_p', `primaryattrition_control_p', `primaryattrition_percent'‘ _n

file close attrition

* Effect size calculation appendix table

cap file close esizecalc

```
file open esizecalc using output/esizecalc.csv, write replace

file write esizecalc ‘ , , , Control group, , Intervention group’ _n

file write esizecalc ‘Outcome, Unadjusted difference in means, Adjusted difference in means, n (missing), Outcome SD, n (missing), Outcome SD, Pooled SD, Hedges' correction (J)’ _n

file write esizecalc ‘WAM Score (ideas double weighted), `wam_treatdiff_unadj', `wam_treatdiff', ‘

file write esizecalc ‘`primarymeasure_control_n' (`primarymeasure_control_miss'), `wam_controlsd', ‘

file write esizecalc ‘`primarymeasure_treat_n' (`primarymeasure_treat_miss'), `wam_treatsd', ‘

file write esizecalc ‘`wam_sdpooled', `wam_j'‘ _n

file write esizecalc ‘WSEM Score, `wsem_treatdiff_unadj', `wsem_treatdiff', ‘

file write esizecalc ‘`wsem_control_n' (`wsem_control_miss'), `wsem_controlsd', ‘

file write esizecalc ‘`wsem_treat_n' (`wsem_treat_miss'), `wsem_treatsd', ‘

file write esizecalc ‘`wsem_sdpooled', `wsem_j'‘ _n

file write esizecalc ‘Ideation Score, `idea_treatdiff_unadj', `idea_treatdiff', ‘

file write esizecalc ‘`idea_control_n' (`idea_control_miss'), `idea_controlsd', ‘

file write esizecalc ‘`idea_treat_n' (`idea_treat_miss'), `idea_treatsd', ‘

file write esizecalc ‘`idea_sdpooled', `idea_j'‘ _n

file close esizecalc


* Power Calculation

powercalc, cluster n(`wam_nhmean') g(`wam_schoolcount') icc(`wam_icc') r2(`wam_r2_pupil') r2_g(`wam_r2_school') treated(`wam_treatprop') blocks(8) regressors(2)

local wam_mdes : di %7.2fc r(mdes)

powercalc, cluster n(`wsem_nhmean') g(`wsem_schoolcount') icc(`wsem_icc') r2(`wsem_r2_pupil') r2_g(`wsem_r2_school') treated(`wsem_treatprop') blocks(8) regressors(2)

local wsem_mdes : di %7.2fc r(mdes)

powercalc, cluster n(`idea_nhmean') g(`idea_schoolcount') icc(`idea_icc') r2(`idea_r2_pupil') r2_g(`idea_r2_school') treated(`idea_treatprop') blocks(8) regressors(2)

local idea_mdes : di %7.2fc r(mdes)


* Power Calculation – FSM sub-group

powercalc, cluster n(`wam_fsm_nhmean') g(`wam_schoolcount') icc(`wam_fsm_icc') r2(`wam_fsm_r2_pupil') r2_g(`wam_fsm_r2_school') treated(`wam_fsm_treatprop') blocks(8) regressors(2)

local wam_fsm_mdes : di %7.2fc r(mdes)

powercalc, cluster n(`wsem_fsm_nhmean') g(`wsem_schoolcount') icc(`wsem_fsm_icc') r2(`wsem_fsm_r2_pupil') r2_g(`wsem_fsm_r2_school') treated(`wsem_fsm_treatprop') blocks(8) regressors(2)

local wsem_fsm_mdes : di %7.2fc r(mdes)

powercalc, cluster n(`idea_fsm_nhmean') g(`idea_schoolcount') icc(`idea_fsm_icc') r2(`idea_fsm_r2_pupil') r2_g(`idea_fsm_r2_school') treated(`idea_fsm_treatprop') blocks(8) regressors(2)

local idea_fsm_mdes : di %7.2fc r(mdes)
```

```
* Table with inputs for power calculation tables

cap file close powercalc

file open powercalc using output/powercalcinputs.csv, write replace

file write powercalc ‘Outcome, MDES, Pupil-level R2, School-level R2, Overall R2, ICC, Average Cluster Size, Control Schools, Intervention Schools, Total Schools, Control Pupils, Intervention Pupils, Total Pupils’ _n

file write powercalc ‘WAM Score, `wam_mdes', `wam_r2_pupil', `wam_r2_school', `wam_r2_overall', `wam_icc', `wam_nhmean',’

    file write powercalc ‘`wam_schoolcount_control', `wam_schoolcount_treat', `wam_schoolcount', `primarysample_control_n', `primarysample_treat_n', `primarysample_n'‘ _n

file write powercalc ‘WSEM Score, `wsem_mdes', `wsem_r2_pupil', `wsem_r2_school', `wsem_r2_overall', `wsem_icc', `wsem_nhmean',’

    file write powercalc ‘`wsem_schoolcount_control', `wsem_schoolcount_treat', `wsem_schoolcount', `secondarysample_control_n', `secondarysample_treat_n', `secondarysample_n'‘ _n

file write powercalc ‘Ideation Score, `idea_mdes', `idea_r2_pupil', `idea_r2_school', `idea_r2_overall', `idea_icc', `idea_nhmean', ‘

    file write powercalc ‘`idea_schoolcount_control', `idea_schoolcount_treat', `idea_schoolcount', `secondarysample_control_n',  `secondarysample_treat_n', `secondarysample_n'‘ _n

file write powercalc ‘WAM Score (FSM), `wam_fsm_mdes', `wam_fsm_r2_pupil', `wam_fsm_r2_school', `wam_fsm_r2_overall', `wam_fsm_icc', `wam_fsm_nhmean',’

    file write powercalc ‘`wam_schoolcount_control', `wam_schoolcount_treat', `wam_schoolcount', `primarysample_fsm_control_n', `primarysample_fsm_treat_n', `primarysample_fsm_n'‘ _n

file close powercalc


* Robustness check including additional controls (time to test, marker fixed effects, KS1)

regress wam_score_dw i.treat FSM EAL FSMgroup EALgroup i.block if primarysample==1 & TestingDate<., vce(cluster anonschoolid)

estimates store notestdate

estadd scalar es = `wam_j'*(_b[1.treat]/`wam_sdpooled')

regress wam_score_dw i.treat FSM EAL FSMgroup EALgroup i.block TestingDate if primarysample==1, vce(cluster anonschoolid)

estimates store testdate

estadd scalar es = `wam_j'*(_b[1.treat]/`wam_sdpooled')


recode wam_marker_num .=0

regress wam_score_dw i.treat FSM EAL FSMgroup EALgroup i.block if primarysample==1 & wam_marker_num<., vce(cluster anonschoolid)

estimates store nomarker

estadd scalar es = `wam_j'*(_b[1.treat]/`wam_sdpooled')

regress wam_score_dw i.treat FSM EAL FSMgroup EALgroup i.block i.wam_marker_num if primarysample==1, vce(cluster anonschoolid)
```

```
estimates store marker
estadd scalar es = `wam_j'*(_b[1.treat]/`wam_sdpooled')
recode wam_marker_num 0=.

mi set flong
mi register imputed tks1ave_round
mi impute regress tks1ave_round FSM EAL FSMgroup EALgroup i.block if primarysample==1, add(20) rseed(11849)
mi estimate, post: regress wam_score_dw i.treat FSM EAL FSMgroup EALgroup i.block if primarysample==1 & tks1ave_round<., vce(cluster anonschoolid)
estimates store noks1
estadd scalar es = `wam_j'*(_b[1.treat]/`wam_sdpooled')
mi estimate, post: regress wam_score_dw i.treat FSM EAL FSMgroup EALgroup i.block tks1ave_round if primarysample==1, vce(cluster anonschoolid)
estimates store ks1
estadd scalar es = `wam_j'*(_b[1.treat]/`wam_sdpooled')
mi extract 0, clear

estimates restore primary
estimates store primary
estadd scalar es = `wam_g'

estout primary notestdate testdate nomarker marker noks1 ks1 using output/robustness.csv, replace ///
   mlabels('Primary' 'No Test Date' 'Test Date' 'No Marker FE' 'Marker FE' 'No KS1' 'KS1', nonumbers nodepvars) ///
   keep(1.treat) varlabels(1.treat 'Treatment') ///
   indicate('Test Date = TestingDate' 'Marker FE = 4.wam_marker_num' 'KS1 = tks1ave_round') ///
   cells(b(fmt(3) star) p(fmt(3) par)) delimiter(', ') ///
   stats(es N, fmt(2 0) labels('Effect Size' 'N'))

* Randomisation inference as a robustness check
ritest treat _b[1.treat], r(2000) strata(block) cluster(anonschoolid) seed(987234) : ///
        regress wam_score_dw i.treat FSM EAL FSMgroup EALgroup i.block, vce(cluster anonschoolid)
local wam_ri_p : di %7.2fc r(p)[1,1]
ritest treat _b[1.treat], r(2000) strata(block) cluster(anonschoolid) seed(987234) : ///
        regress wsem16_score i.treat FSM EAL FSMgroup EALgroup i.block, vce(cluster anonschoolid)
local wsem_ri_p : di %7.2fc r(p)[1,1]
ritest treat _b[1.treat], r(2000) strata(block) cluster(anonschoolid) seed(987234) : ///
        regress ideation_score i.treat FSM EAL FSMgroup EALgroup i.block, vce(cluster anonschoolid)
local idea_ri_p : di %7.2fc r(p)[1,1]
```

```
* Table to compare p values from randomisation inference and sampling inference
cap file close ricomp
file open ricomp using output/ricomp.csv, write replace
file write ricomp ‘Outcome, Effect size, Sampling Inference p-value, Randomisation Inference p-value’ _n
file write ricomp ‘WAM Score (ideas scale double weighted), `wam_g', `wam_treatdiff_p', `wam_ri_p'‘ _n
file write ricomp ‘WSEM Score, `wsem_g', `wsem_treatdiff_p', `wsem_ri_p'‘ _n
file write ricomp ‘Ideation Score, `idea_g', `idea_treatdiff_p', `idea_ri_p'‘ _n
file close ricomp
```

Appendix G: Teacher interview guide

Power of Pictures: Interviews with Classroom Teacher

The interviews should last around 30 minutes. The timings given for each section are a guide – you may spend longer or shorter on each section. Lead questions are presented in bold, with potential follow-up questions presented in a non-bold typeface. As the interviews are semi-structured, not all questions need to be asked and they do not need to be asked in order. The interviewer should be responsive to what the interviewee, following the direction of the conversation and following-up with additional questions as needed.

Main objective	Purpose of section	Guide timings
1. Introduction	Explains the purpose and ‘ground rules’ of the interview.	3 mins
2. Background context	Allows the participant an opportunity to settle into the interview, as well as providing some background to the school, the teacher’s prior experience and the context in which the Power of Pictures programme is being delivered.	5 mins
3. Experience of CPD (Continuing Professional Development)/ Training	This section will focus on understanding the perceived quality of the CPD/training, and how teachers translated the training into practice in their classroom. Through this section we hope to gain insight into differences in the delivery and experience of the programme, depending on which author is co-delivering the programme.	5 mins
4. Experience of the author workshop	This section will focus on understanding the perceived quality and purpose of the author workshop, as well as children’s engagement with it. Through this section we hope to gain insight into differences in the delivery and experience of the programme, depending on which author is co-delivering the programme.	2 mins
5. Delivery of the Power of Pictures lessons	To understand the teacher’s experience of delivering the PoP lessons, including barriers and facilitators to delivery, particularly considering SLT buy-in and timetabling of the lessons. This section will help us to understand how different teachers implement the problem, and also help us to understand how Power of Pictures is different to the usual teaching/curriculum.	5 mins
6. Programme impact and mechanisms	To explore the teacher’s perception of pupils’ engagement with Power of Pictures and the positive and negative impact of the programme for the teacher and pupils, together with the mechanisms that brought about any impact identified.	5 mins

7. Close	Thank you and close	2 mins
Observation of PoP session		
Reflection following observation of PoP lesson	If PoP lessons are ongoing (some schools will have completed their PoP lessons), this will be an opportunity to follow-up on any areas of interest arising from the observation.	3 mins

1. Introduction	3 mins
Introduction: ● Introduce yourself ● Introduce BIT and IOE – explain that we are independently evaluating the Power of Pictures programme, which is one of five programmes that are part of the Cultural Learning programme that is jointly funded by the Education Endowment Foundation and the Royal Society of Arts. Aims of this interview: We are here to learn more about how the Power of Pictures programme has worked in your class. We're interested in what involvement you have had with the programme and what has helped the programme to work, and what the challenges have been. We'd also like to understand any impact the programme has had on your school, particularly pupils in your class. This interview: ● Should take around 30 minutes ● Stress that you want to understand the intervention from their point of view. No answers are right or wrong – and we are not here to judge the decisions made or views held by the interviewee. Anonymity and privacy: ● All information gathered will be in strict confidence, unless there are concerns about safeguarding. When we write up the research we will ensure that no one is identifiable from any reporting. ● Explain that if at any point they feel uncomfortable or prefer not to answer a specific question they can just say so. ● Explain that it is their choice whether they take part in the interview and they can end the interview at any point, without giving a reason.	Orientates respondent and gets them prepared to take part in the discussion. Outlines the 'rules' of the interview.

Recording: 1. Explain that recording enables us to have an accurate record of what was said, which can be typed up for analysis alongside other interviews. We may also use quotes from this interview, but these will be included in a way that means no individual or school is identifiable. 2. Check if they have any questions about the interview. If they are happy to go ahead, obtain verbal permission to digitally record and take notes (written permission should already have been obtained). 3. Once you have consent, start the voice recorder. 4. State interview number/participant ID	
2. Background context	5 mins
How many years have you been teaching? How long have you been working at this school? What would you say are the school's main strengths and challenges? ● What access do children at this school typically have to books? (Probe for both at home and at school). Before you took part in Power of Pictures, what experience did you have of: a) Creative writing b) Art and drawing c) Teaching using picturebooks How did you become involved with the Power of Pictures programme? ● What were your thoughts about the programme when you first heard of it?	Allows the participant an opportunity to settle into the interview, as well as providing some background to the school and the context in which the Power of Pictures programme is being delivered.
3. Experience of CPD/training	5 mins
Can you tell me a bit about the CPD/training you did as part of Power of Pictures? ● What did you think of the CPD/training? ● Can you tell me about anything you found helpful about the CPD/training? ● Can you tell me about anything you would change about the CPD/training? ● What did you think of the way the facilitators delivered the CPD/training?	This section will focus on understanding the perceived quality of the training and how teachers translated the training into practice in their classroom.

• Following the CPD/training, how did you feel about delivering the Power of Pictures lessons with your class? Can you tell me about the author/illustrator's role in the training? • What did you think of the training they delivered? • How did it compare to the rest of the CPD training you received? How did you find translating the CPD training into your teaching practice? • Which elements of the CPD have you used in your teaching? • How did you find that worked with your class? • Can you tell me about any aspects of the training that you did not use in the classroom? • Why was this?	
4. Experience of author workshop	2 mins
Can you tell me about the author workshop you attended with your class as part of Power of Pictures? • How did the workshop fit within the wider Power of Pictures programme? • What did you think of the workshop? • How did the children respond to the workshop? • What do you think worked well about the workshop? • Can you tell me about anything that you'd change about the workshop?	
5. Delivery of the Power of Pictures lessons	5 mins
Have you ever used picturebooks to teach before you began Power of Pictures? • If yes: Can you tell me about what you did? Probe to find out if this was Key Stage 2 teaching. • If no: Why do you think you hadn't done so? • How does the Powers of Pictures programme compare to how you would usually teach writing and drawing to Year 5? Can you tell me about what you've covered in the Power of Pictures lessons that you've delivered to your class? • How did you find delivering the Power of Pictures lessons? • What do you think went well? • What do you think didn't work so well? • What recommendations do you have as to how the programme could be improved?	

How were the Power of Pictures lessons organised and timetabled in your curriculum? ● When in the school year did you timetable them? Why was this? ● What content would have otherwise been covered in the curriculum? ● What arrangements did you make in order for you to attend the CPD training? ● How did this work out? How have the SLT been involved in the delivery of the Power of Pictures programme? ● Can you describe any support you've received from SLT to deliver the programme? ● Can you describe any resources they've made available to deliver the programme?	
6. Programme Impact and Mechanisms	5 mins
How have pupils responded to the Power of Pictures lessons? ● To what extent have pupils engaged with the sessions? ● Can you describe any factors that have affected children's engagement with Power of Pictures? What effect do you think taking part in Power of Pictures has had for pupils? ● What impact, if any, has the programme had on children's writing? ● What impact if any, has the programme had on children's drawing? ● What was it about the programme that you think led to [change described]? ● Can you describe any pupils that you think the programme is particularly helpful for? Why do you think that is? ● Can you describe any pupils who you think have got less out of the programme? Why do you think that is? ● Have there been any negative consequences of the programme for pupils? If so, can you describe what they are. What effect, if any, do you think that the programme has had for you? ● Probe: attitude to teaching with picturebooks in KS2; confidence teaching writing and drawing. ● What do you think it was about the programme that led to these changes? Can you describe any ways that you have shared learning from the programme beyond your class?	To explore the teacher's perception of pupils' engagement with Power of Pictures and the positive and negative impact of the programme, together with the mechanisms that brought about any impact identified.

• What effect, if any, do you think the programme has had on the wider school? Probe impact on: pupils in other classes; other teachers.	
7. Close	2 mins
Overall, would you recommend Power of Pictures to other schools? • Why/Why not? Was there anything else that you were hoping to discuss that we haven't yet had a chance to talk about? Thank the interviewee for their time and reassure them of the confidentiality of their responses, as explained at the beginning of the interview.	Thank you and close.

Reflection following observation of PoP session	5 mins
Below are some exemplar questions, however, the interviewer may want to add some additional questions based on their observations. How did you find that session went? • What do you think went well? • What do you think the challenges were? • Which students engaged particularly well? How/why?	Opportunity to follow-up on any areas of interest arising from the observation.

• Which students engaged less well? How/why? • If you could go back and do that session again, what might you do differently? How does that session compare to other Power of Pictures lessons you've taught? Probe specific areas of interest from your observation/in relation to the evaluation questions/thinking about how this school is running Power of Pictures in comparison to other schools. Could you describe any ways that the sessions have changed over time?	

Appendix H: Senior leadership team (SLT) interview guide

Power of Pictures: Interviews with Senior Leadership Team

The interviews should last around 30 minutes. The timings given for each section are a guide – you may spend longer or shorter on each section. Lead questions are presented in bold, with potential follow-up questions presented in a non-bold typeface. As the interviews are semi-structured, not all questions need to be asked and they do not need to be asked in order. The interviewer should be responsive to what the interviewee, following the direction of the conversation and following-up with additional questions as needed.

Main objective	Purpose of section	Guide timings
1. Introduction	Explains the purpose and 'ground rules' of the interview.	3 mins
2. Background context	Allows the participant an opportunity to settle into the interview, as well as providing some background to the school, so that we understand more about the context in which the PoP programme is being delivered.	4 mins
3. Engagement	To understand the extent to which the school, and particularly the SLT, has engaged with and supported the programme.	7 mins
4. Delivery experience	This section will focus on understanding the perceived quality of the intervention, as well as the experience of the programme's delivery, including barriers and facilitators to delivery. We will also try and disentangle what is 'business as normal' within the school.	7 mins
5. Mechanisms of change	To explore the SLT member's perception of the positive and negative impact of the programme, particularly for pupils, and the mechanisms that brought about any change identified.	7 mins
6. Close	Thank you and close	2 mins

1. Introduction	3 mins
Introduction: ● Introduce yourself ● Introduce BIT and IOE – explain that we are independently evaluating the Power of Pictures programme, which is one of five programmes that are part of the Cultural Learning programme that is jointly funded by the Education Endowment Foundation and the Royal Society of Arts. Aims of this interview: We are here to learn more about how the PoP programme has worked in your school. We're interested in what involvement you have had with the programme, what has helped the programme to work, and what the challenges have been. We'd also like to understand any impact the programme has had on your school, particularly pupils in the PoP class. This interview: ● Should take no more than 30 minutes ● Stress that you want to understand the intervention from their point of view. No answers are right or wrong – and we are not here to judge the decisions made or views held by the interviewee. Anonymity and privacy: ● All information gathered will be in strict confidence, unless there are concerns about safeguarding. When we write up the research we will ensure that no one is identifiable from any reporting. ● Explain that if at any point they feel uncomfortable or prefer not to answer a specific question they can just say so. ● Explain that it is their choice whether they take part in the interview and they can end the interview at any point, without giving a reason. Recording: 1. Explain that recording enables us to have an accurate record of what was said, which can be typed up for analysis alongside other interviews. We may also use quotes from this interview, but these will be included in a way that means no individual or school is identifiable. 2. Check if they have any questions about the interview. If they are happy to go ahead, obtain verbal permission to digitally record and take notes (written permission should already have been obtained). 3. Once you have consent, start the voice recorder.	Orientates respondent and gets them prepared to take part in the discussion. Outlines the 'rules' of the interview.

4. State interview number/participant ID	
2. Background context	4 mins
Can you start by telling me how long you've been working at the school and what your role is? Could you tell me a bit about the school? ● What would you say are the schools' main strengths? ● What would you say are some of the school's biggest challenges? ● In terms of reading, what access do children at the school have to books at school? ● What access do children have to book in their home environment? How did your school become involved with the Power of Pictures programme? ● When did you first hear about the Power of Pictures programme? ● Why did your school decide to get involved in the programme? ● What initial expectations did you have for the programme?	Allows the participant an opportunity to settle into the interview, as well as providing some background to the school, so that we understand more about the context in which the PoP programme is being delivered.
3. Engagement	7 mins
Can you describe what your involvement with the Power of Pictures programme has been? ● Can you describe any support that you or other members of SLT have provided to the programme? ● Can you describe any resources that you've made available to enable the programme to run? Have you sat in on any of the Power of Pictures lessons? If yes: ● Can you describe any sessions you've sat in on? ● What did you think of the session(s)? [probe for further exploration of strength and suggested areas for improvement] If no: ● Were there any particular reasons that you didn't sit in on a session?	To understand the extent to which the school, and particularly the SLT, has engaged with and supported the programme

• What do you understand the class have been doing in their Power of Pictures sessions?	
4. Delivery experience	7 mins
Overall, what do you think of the quality of the Powers of Pictures programme? • What do you think has worked well about the programme? • What have been the main challenges of the programme? • What have you done to try and overcome these challenges? • Knowing what you know now, is there anything you would approach differently about Power of Pictures? How have you found organising the Power of Pictures lessons in the school timetable? • What content would otherwise have been covered in the curriculum? • What arrangements did you make in order for the class teacher to attend the CPD training? • How did this work out? How is writing typically taught to pupils in Year 5 at your school? • What opportunities do teachers typically give to children to write in Year 5? • What books do you typically use when teaching Year 5? • Would children previously have been taught using pictures books in Key Stage 2? • How does the Power of Pictures programme compare to the school's business-as-usual literacy teaching?	This section will focus on understanding the perceived quality of the intervention, as well as the experience of the programme's delivery, including barriers and facilitators to delivery. We will also try and disentangle what is 'business as normal' within the school.
5. Mechanisms of change	7 mins
From what you've heard or seen directly, how have pupils responded to the Power of Pictures programme? • To what extent have pupils engaged with the sessions? • Can you describe any factors that affect children's engagement with the programme? What effect do you think taking part in Power of Pictures has had for pupils? • What impact, if any, has the programme had on children's writing? • What impact if any, has the programme had on children's drawing? • What was it about the programme that you think led to [change described]?	To explore the SLT member's perception of the positive and negative impact of the programme, particularly for pupils, and the mechanisms that brought about any change identified.

• Can you describe any pupils that you think the programme is particularly helpful for? • Can you describe any pupils who you think have got less out of the programme? • Can you describe any negative impacts of the programme? What effect, if any, do you think that the programme has had for [name of class teacher]? • Can you describe any impact that the programme has had on their teaching practice? • Probe: confidence writing/drawing; attitude to books; attitude to teaching with picturebooks Can you describe any ways that the programme has been shared within the wider school? • What effect, if any, do you think the programme has had more broadly within the school? Probe: pupils in other classes; other teachers.	
6. Close	2 mins
Overall, would you recommend Power of Pictures to other schools? • Why/Why not? Was there anything else that you were hoping to discuss that we haven't yet had a chance to talk about? Thank the interviewee for their time and reassure them of the confidentiality of their responses, as explained at the beginning of the interview.	Thank you and close

Appendix I: Guide for information discussions with pupils

Power of Pictures: Discussions with Pupils

Short discussions, lasting around five minutes will be conducted with pupils happy to speak to a researcher. These will happen during lesson time in the classroom setting. The researcher(s) will take fieldnotes and therefore the conversations will not be audio recorded.

Aims of discussions with pupils

To better understand pupils' responses to and engagement with the Power of Pictures programme.

Before conducting discussions

- The researcher should find out from the teacher in advance which pupils' parents have consented for them to take part in the research, and whether there are pupils who have particular communication needs, where the researchers will need to adapt their approach.
- The teacher should explain to the class: who the researcher is and what they are going to be doing; the purpose of the research; and reassure pupils that they do not have to talk to the researcher if they would prefer not to.
- Before speaking to a pupil, ask whether they are okay to speak to you and ensure that their body language indicates that they are happy to talk to you.

Discussion guide questions

Can you tell me about what you do/did in your Power of Pictures lessons?

- Can you tell me about some of the things you've learnt?
- Can you tell me about anything you've liked about them?
- Can you tell me about anything you haven't liked?
- Is there anything you would change about the Power of Pictures lessons?

Can you describe the picturebook you made/you're making to me?

- Where did you get the ideas for your picturebook?
- How did you find making it?
- (If complete) What did you think of the final picturebook you made?
- Did you show anyone your picturebook? If yes, what did they think?

(Ideally they may be able to show you the picturebook and you can ask specific questions about how they wrote/drew certain things, why they included certain characters etc.)

Can you tell me about the day when you went to meet the author/illustrator?

- What did you do with them?
- How did you find it?
- What did you think of the author/illustrator?

Appendix J: Observation proforma

Power of Pictures: Observation proforma

This proforma may be completed during the observation, or more extensive notes may be made during the observation, i.e. in a notepad, and then summarised on this form.

Date and time: Name of school and class: Number of pupils present: Name of observer: Other adults present: Notes	
Description of setting, space and resources used for the lesson, e.g. type of room and how desks are arranged; types of drawing/art materials available.	
What is covered in the introduction to the session, e.g. are the aims and expectations covered?	
What activities are carried out as part of the session, e.g. nature of individual, small group activities and/or whole class activities?	

What types of work do children produce as part of the lesson?	
How is feedback provided to children throughout the lesson?	
How confident is the teacher in delivering the lesson?	
To what extent are children engaged in the session, e.g. are they focused on writing and drawing, or are they distracted and talking to their peers about unrelated things?	
What is the behaviour of the pupils like, e.g. mainly on-task; mainly disruptive; individual children not on-task but the majority engaged?	
To what extent do children seem able to understand and follow the session? What difficulties do children who struggle seem to have? How are they supported?	
What strategies are used to account for the needs and abilities of different pupils?	

Other	

Appendix K: Raw treatment condition survey data

Please note: some percentages may add up to slightly above or below 100% due to rounding.

1.1 What is your role at the school?	
Response	Number (%)
Class teacher	42 (98%)
Other members of SLT (e.g. Head of Key Stage 2)	4 (9%)
Deputy headteacher	1 (2%)
Other (please specify): - Science coordinator - PPA release teacher - Year 5 leader	3 (7%)
Headteacher	1 (2%)

N.B. Total percentage is greater than 100% because respondents could select more than one role.

2.1 To what extent do you agree or disagree with the following statement: Children at this school are able to access a range of books in their homes	
Response	Number (%)
Agree	13 (30%)
Neither agree nor disagree	13 (30%)
Disagree	11 (26%)
Strongly agree	4 (9%)
Strongly disagree	2 (5%)

2.2 To what extent do you agree or disagree with the following statement: Children at this school are able to access a range of books at school	
Response	Number (%)
Strongly agree	25 (58%)
Agree	17 (40%)
Disagree	1 (2%)
Neither agree nor disagree	0 (0%)
Strongly disagree	0 (0%)

3.1 How many training sessions did you attend as part of the Power of Pictures programme?	
Response	Number (%)
Three	36 (84%)
Two	5 (12%)
One	1 (2%)
Don’t know	1 (2%)

N.B. Respondents answering ‘three’ to question 3.1 were not asked to complete question 3.2

3.2 Which of the following, if any, are reasons for your limited attendance at Power of Pictures training?	
Response	Number (%)
I could not find cover	3 (7%)
I did not know about the training	1 (2%)
I was unwell	1 (2%)

I had competing school commitments	1 (2%)
Other (please specify): - recorded wrong date in diary	1 (2%)
Cost of travel and/or subsistence	0 (0%)
I could not travel the distance required	0 (0%)
I did not feel my my attendance would be useful	0 (0%)
I did not receive time off in lieu	0 (0%)
I had personal commitments	0 (0%)
SLT were not supportive	0 (0%)

N.B. Respondents could select more than one answer.

3.3 After you received your training, how prepared did you feel to deliver the Power of Pictures lessons to your class?

Response	Number (%)
I felt very prepared	32 (76%)
I felt somewhat prepared	10 (24%)
I felt somewhat unprepared	0 (0%)
I did not feel prepared at all	0 (0%)
Don’t know	0 (0%)

N.B. The respondent answering ‘Don’t know’ to 3.1 was not asked to complete questions 3.3–3.5.

3.4 Thinking about the facilitator who delivered your training, to what extent do you agree or disagree with the following statements: The facilitator was good at explaining things to me

Response	Number (%)
Strongly agree	34 (81%)
Agree	7 (17%)
Neither agree nor disagree	1 (2%)
Disagree	0 (0%)
Strongly disagree	0 (0%)

N.B. The respondent answering ‘Don’t know’ to 3.1 was not asked to complete questions 3.3–3.5.

3.5 Thinking about the facilitator who delivered your training, to what extent do you agree or disagree with the following statements: I found the facilitator to be engaging	
Response	Number (%)
Strongly agree	35 (83%)
Agree	6 (14%)
Neither agree nor disagree	1 (2%)
Disagree	0 (0%)
Strongly disagree	0 (0%)

N.B. The respondent answering ‘Don’t know’ to 3.1 was not asked to complete questions 3.3–3.5.

4.1 As part of Power of Pictures, you worked with an author/illustrator. Who did you work with?	
Response	Number (%)
Viviane Schwarz	20 (47%)
Alexis Deacon	18 (42%)
Tim Hopgood	5 (12%)

4.2 Thinking about the author/illustrator you worked with, to what extent do you agree or disagree with the following statements: I found the author/illustrator to be engaging	
Response	Number (%)
Strongly agree	31 (72%)
Agree	12 (28%)
Neither agree nor disagree	0 (0%)
Disagree	0 (0%)
Strongly disagree	0 (0%)

4.3 Thinking about the author/illustrator you worked with, to what extent do you agree or disagree with the following statements: The pupils found the author/illustrator to be engaging	
Response	Number (%)
Strongly agree	31 (72%)
Agree	12 (28%)
Neither agree nor disagree	0 (0%)
Disagree	0 (0%)
Strongly disagree	0 (0%)

4.4 Thinking about the author/illustrator you worked with, to what extent do you agree or disagree with the following statements: The author/illustrator was good at explaining things to me	
Response	Number (%)
Strongly agree	32 (74%)

Agree	9 (21%)
Neither agree nor disagree	2 (5%)
Disagree	0 (0%)
Strongly disagree	0 (0%)

4.5 Thinking about the author/illustrator you worked with, to what extent do you agree or disagree with the following statements: The author/illustrator was good at explaining things to the pupils

Response	Number (%)
Strongly agree	28 (65%)
Agree	12 (28%)
Neither agree nor disagree	3 (7%)
Disagree	0 (0%)
Strongly disagree	0 (0%)

4.6 Thinking about the author/illustrator you worked with, to what extent do you agree or disagree with the following statements: The author/illustrator motivated the pupils to produce their picturebook

Response	Number (%)
Strongly agree	31 (72%)
Agree	9 (21%)
Neither agree nor disagree	3 (7%)
Disagree	0 (0%)
Strongly disagree	0 (0%)

5.1 How engaging, if at all, did you think your Power of Pictures focus picturebook was for your pupils?	
Response	Number (%)
Very engaging	31 (72%)
Somewhat engaging	11 (26%)
Neither engaging nor unengaging	1 (2%)
Somewhat unengaging	0 (0%)
Very unengaging	0 (0%)

6.1 How confident, if at all, do you feel about the following: I understand the purpose of the Power of Pictures programme	
Response	Number (%)
Very confident	35 (81%)
Somewhat confident	8 (18%)
Neither confident or not confident	0 (0%)
Somewhat unconfident	0 (0%)
Not confident at all	0 (0%)

6.2 How confident, if at all, do you feel about the following: I understand the features of the Power of Pictures programme	
Response	Number (%)
Very confident	27 (63%)
Somewhat confident	14 (33%)

Neither confident or not confident	2 (5%)
Somewhat unconfident	0 (0%)
Not confident at all	0 (0%)

6.3 How confident, if at all, do you feel about the following: I understand how picturebooks are created

Response	Number (%)
Very confident	34 (79%)
Somewhat confident	7 (16%)
Neither confident or not confident	2 (5%)
Somewhat unconfident	0 (0%)
Not confident at all	0 (0%)

6.4 How confident, if at all, do you feel about the following: I can integrate techniques from the Power of Pictures training in the classroom

Response	Number (%)
Somewhat confident	31 (72%)
Very confident	10 (23%)
Neither confident or not confident	2 (5%)
Somewhat unconfident	0 (0%)
Not confident at all	0 (0%)

7.1 Thinking about your class as a whole, what kind of impact, if any, do you think your Power of Pictures experience has had on: pupils' ideation (ability to come up with ideas)

Response	Number (%)
Very positive impact	22 (51%)
Somewhat positive impact	19 (44%)
No impact	2 (5%)
Somewhat negative impact	0 (0%)
Very negative impact	0 (0%)
Don’t know	0 (0%)

7.2 Thinking about your class as a whole, what kind of impact, if any, do you think your Power of Pictures experience has had on: pupils’ behaviour

Response	Number (%)
Very positive impact	14 (33%)
Somewhat positive impact	8 (19%)
No impact	20 (47%)
Somewhat negative impact	0 (0%)
Very negative impact	0 (0%)
Don’t know	1 (2%)

7.3 Thinking about your class as a whole, what kind of impact, if any, do you think your Power of Pictures experience has had on: pupils’ reading

Response	Number (%)
Very positive impact	10 (23%)

Somewhat positive impact	27 (63%)
No impact	6 (14%)
Somewhat negative impact	0 (0%)
Very negative impact	0 (0%)
Don’t know	0 (0%)

7.4 Thinking about your class as a whole, what kind of impact, if any, do you think your Power of Pictures experience has had on: pupils’ writing skills

Response	Number (%)
Very positive impact	10 (23%)
Somewhat positive impact	26 (60%)
No impact	4 (9%)
Somewhat negative impact	3 (7%)
Very negative impact	0 (0%)
Don’t know	0 (0%)

7.5 Thinking about your class as a whole, what kind of impact, if any, do you think your Power of Pictures experience has had on: pupils’ communication skills

Response	Number (%)
Very positive impact	12 (28%)
Somewhat positive impact	23 (53%)
No impact	8 (19%)

Somewhat negative impact	0 (0%)
Very negative impact	0 (0%)
Don’t know	0 (0%)

7.6 Thinking about your class as a whole, what kind of impact, if any, do you think your Power of Pictures experience has had on: pupils’ social skills

Response	Number (%)
Very positive impact	8 (19%)
Somewhat positive impact	17 (40%)
No impact	17 (40%)
Somewhat negative impact	0 (0%)
Very negative impact	0 (0%)
Don’t know	1 (2%)

7.7 Thinking about your class as a whole, what kind of impact, if any, do you think your Power of Pictures experience has had on: pupils’ creativity

Response	Number (%)
Very positive impact	29 (67%)
Somewhat positive impact	13 (30%)
No impact	1 (2%)
Somewhat negative impact	0 (0%)
Very negative impact	0 (0%)

Don’t know	0 (0%)

7.8 Thinking about your class as a whole, what kind of impact, if any, do you think your Power of Pictures experience has had on: pupils’ engagement	
Response	Number (%)
Very positive impact	30 (70%)
Somewhat positive impact	12 (28%)
No impact	1 (2%)
Somewhat negative impact	0 (0%)
Very negative impact	0 (0%)
Don’t know	0 (0%)

7.8 Thinking about your class as a whole, what kind of impact, if any, do you think your Power of Pictures experience has had on: pupils’ confidence	
Response	Number (%)
Very positive impact	23 (53%)
Somewhat positive impact	18 (42%)
No impact	2 (5%)
Somewhat negative impact	0 (0%)
Very negative impact	0 (0%)
Don’t know	0 (0%)

8.1 Overall, how would you rate Power of Pictures?	
Response	Number (%)
Very good	29 (68%)
Good	11 (26%)
OK	3 (7%)
Poor	0 (0%)
Very poor	0 (0%)
Don’t know	0 (0%)

Appendix L: Raw control condition survey data

Please note: some percentages may add up to slightly above or below 100% due to rounding.

1.1 What is your role at the school?	
Response	Number (%)
Class teacher	34 (83%)
Other members of SLT	3 (7%)
Head or Deputy Head	2 (5%)
Other - Year 5 teacher, literacy leader - Year 5 teacher, Year 5 lead, English lead	 1 (2%) 1 (2%)

2.1 Have you ever received training on using picturebooks in Key Stage 2 teaching?	
Response	Number (%)
No	31 (76%)
Yes	10 (24%)

2.2 When did you receive this training? (For participants answering ‘Yes’ to 2.1)	
Response	Number (%)
Within the last academic year	6 (15%)
Between 2 and 5 years ago	3 (7%)
More than 5 years ago	1 (2%)

3.1 Over the last academic year, how often have you used picturebooks to teach your class?	
Response	Number (%)
A few times	24 (59%)
Never	7 (17%)
Monthly	6 (15%)
Every couple of weeks	2 (5%)
Weekly	2 (5%)

3.2 How confident do you feel teaching your class using picturebooks (where 1 = not confident at all and 5 = very confident)?	
Response	Number (%)
4	15 (37%)
3	14 (34%)
5	6 (15%)
1	3 (7%)
2	3 (7%)

Appendix M: Memorandum of Understanding for schools

The Power of Pictures Evaluation

MEMORANDUM OF UNDERSTANDING

This project is exploring how the opportunity for teachers to work with professional writers might change their understanding of being a writer, how they teach writing, and improve outcomes in writing for the children they teach. Its impact will be evaluated by comparing it with the ‘teaching as usual’ approach using a randomised controlled trial (RCT).

During this project, you will be contacted by both the **Project Team** (Centre for Literacy in Primary Education), who are responsible for developing and supporting the new teaching approach, and by the **Evaluation Team** (University College London [UCL] and Behavioural Insights Team [BIT]), who are carrying out an independent evaluation of its effectiveness.

This memorandum of understanding (MoU) explains what your school's participation in the study will entail.

Randomised Controlled Trial (June 2018–July 2019)

The trial will involve your school being randomly assigned either to participate in the Power of Pictures intervention (the intervention group) or to continue with your normal teaching approach (the comparison group).

Teachers in the intervention group will receive two and a half days of training for a Year 5 teacher to explore specific techniques around using and creating picturebooks in the classroom and a half day whole class workshop with a published author/illustrator to extend children's skills in creating and developing characters in their own narrative writing. Teachers will also receive a copy of a high quality text written and illustrated by their author tutor as well as specially written teaching sequences to implement the training in their own classrooms. Schools in the intervention group will be asked to pay £200 to participate as a partial contribution to the costs of the intervention, and will receive a minimum of £500 towards supply costs.

Teachers in the Comparison group will teach writing as they normally do. Schools in the comparison group will receive a £500 payment, which can be used towards the cost of being involved in the Power of Pictures after the project has ended, if desired.

The following information and evaluation data will be required by the Evaluation and Project teams:

Prior to randomisation

Schools will:

- Provide school URN and LAESTAB number.
- Provide contact details of the Year 5 Project Teacher (valid email address and telephone number) to the Project Team for use by both the Project Team and the Evaluation team.
- Provide, via the means specified by the evaluation team, pupil names, DOB and Unique Pupil Numbers (UPNs) of the Year 5 Project Class, along with details of any setting or streaming by attainment, to the Evaluation Team by the end of March 2018.
- Follow the secure procedures requested by Evaluation team to allow parents to opt their pupils out of the evaluation part of this research.

During the evaluation

Participating teachers will:

- Update UPNs and pupil names of Year 5 Project Class by the end of September 2018.
- Facilitate visits by the Project Team and/or the Evaluation Team to gather data on the implementation of the intervention or on routine teaching (for example, through observations or interviews etc.).
- Facilitate an end-of-project written assessment which will be administered by the Evaluation Team.

Use of Data

All pupil data will be treated with the strictest confidence and will be stored in accordance with the Data Protection Act (1998) and with the forthcoming General Data Protection Regulation. Named data will be matched with the National Pupil Database using pupils' UPNs by the Evaluation Team and shared (anonymously) with the Education Endowment Foundation. All results will be anonymised so that no schools will be identifiable in the report or dissemination of any results.

For the purpose of research, the pupil data will be linked with information from the National Pupil Database held by the Department for Education, other official records, and shared with the Department for Education, EEF, EEF's data contractor FFT Education and in an anonymised form to the UK Data Archive and for research purposes. Confidentiality will be maintained at all times.

Requirements for Schools

- The school is not participating in another research project or evaluation that would interfere with development and evaluation of the above approach in Year 5 writing.
- The Year 5 Project Teacher will be working with a Year 5 class in the academic year 2018-2019.
- Intervention group teachers will attend 2.5 training days, use the teaching sequence as planning for whole class literacy sessions (for a total of four weeks) in between the training days, bring their Year 5 class and the required number of accompanying adults to a 75-minute workshop led by the author/illustrator and bring finished examples of children's own writing to the last half day training session.
- The school will deliver letters to parents giving them information about the study and an opportunity to opt their child out of the data gathering process. They will inform the Evaluation Team of any responses arising.
- The school will provide data requested to the Project Team and the Evaluation Team as detailed above. The school will permit the publication of anonymised data collected and its use in presentations.
- Teachers will, at the earliest opportunity, notify the Project Team if there are support or operational issues which could affect their participation.
- Teachers will complete a survey for each child in their class at the beginning and end of the study.
- If the school has to withdraw from the project for operational or other unavoidable reasons, it will notify the UCL/BIT Evaluation Team straight away and, wherever possible, still provide test data for the project.

Responsibilities of the Project Team:

- Provide 2.5 days of training for Year 5 teachers, co-led by an author/illustrator
- Provide all teachers with a copy of the focus book, a teaching sequence to implement as their whole class literacy teaching and supporting web materials to enable teachers to implement the project effectively in their own classrooms
- Provide and facilitate workshops for each Year 5 class involved led by the author/illustrator
- Act as the first point of contact for any questions about the project
- Provide on-going support to the school
- Provide information sheets for parents
- Collect Year 5 Project Teacher contact names and email details.
- Disseminate the research findings.

Responsibilities of the Evaluation Team:

- Conduct the random allocation
- Work with the project team and schools to collect class and pupil level data (including name, date of birth, UPN) Collect data about how the schools are implementing the intervention, or teaching as usual (control groups)
- Request NPD data using pupil details

- Administer the writing assessment
- Analyse the data from the project
- Disseminate the research findings

Appendix N: WAM Prompt

Adapted from Dunsmuir et al. (2015)

Your name: _____

Imagine that you are going to go on a school trip with your class, your teacher, and some other adults. Imagine you could go anywhere at all.

Write a short story about your chosen trip, and some of the things that you did while on the trip.

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

Appendix O: WSEM Prompt

Adapted from Bruning et al. (2013)

Thinking about myself as a Writer

Name: _____

Instructions: Put a tick in one box for each question below to show how confident YOU feel about the different things a writer does. This isn't a test. We want to know how you really feel when you are writing.

	I'm sure I can't do it	I don't think I can do it	I'm not sure if I can do it or not	I can mostly do it	I'm very sure I can do it
1. I can think of lots of new ideas for my writing.					
2. I can use my ideas in my writing.					
3. I can think of the words I need to write down my ideas.					
4. I can show how I feel in my writing					
5. I know where to place my ideas in my writing.					
6. I can spell my words correctly.					
7. I can write whole sentences.					
Please turn over to the other side					

	I'm sure I can't do it	I don't think I can do it	I'm not sure if I can do it or not	I can mostly do it	I'm very sure I can do it
8. I can use punctuation marks (like full stops and question marks) correctly.					
9. I can begin my paragraphs in the right places.					
10. I can focus on my writing for at least half an hour.					
11. I can start writing quickly.					
12. I am always calm and in control when I write.					
13. I can think of my goals for my writing before I write.					
14. I can keep on writing even when it gets difficult.					
15. I enjoy writing.					
16. I am confident as a writer					

Thank you for helping us by thinking about writing.

Appendix P: School and parent information sheets, objection forms, and revised GDPR privacy notice

School information sheet

The EEF/RSA 'Learning about Culture' Trial of

The Power of Pictures

What is this about?

The Centre for Literacy in Primary Education are currently taking part in a project funded by EEF and the RSA. This aims to improve our understanding of interventions to improve pupils writing skills, including our project the Power of Pictures. The effectiveness of the project will be evaluated by a team from UCL Institute of Education and the Behavioural Insights Team. **This is a great opportunity to help us understand the impact of the Power of Pictures and receive training that helps primary school teachers to develop their understanding of the craft of picturebook creation and illustration as a way of raising children's achievement in literacy.**

What will the project look like?

The Power of Pictures was devised by CLPE expert teachers in partnership with illustrator Ed Vere. The training supports primary teachers to use quality picturebooks to improve all children's critical thinking and writing ability. The project includes two and a half days of training for a Year 5 teacher to explore specific techniques around using and creating picturebooks in the classroom and a half day whole class workshop with a published author/illustrator to extend children's skills in creating and developing characters in their own narrative writing. Teachers will also receive a copy of a high quality text written and illustrated by their author tutor as well as specially written teaching sequences to implement the training in their own classrooms.

What are 'intervention' and 'comparison' schools?

An important element of these EEF projects is that schools interested in participating are randomly allocated to either the intervention group, who will receive the training, or a comparison group, who contribute to the data required for comparison. It is important that all interested schools understand that they could be in either group. ***It is important that you understand you could be a comparison school.***

What commitment would this project require?

If you are an 'intervention' school, you would need to commit to allowing the Year 5 project teacher to:

- provide student profile and attainment data and UPNs;
- attend 2.5 training days
- use the teaching sequence as planning for whole class literacy sessions (for a total of four weeks) in between the training days
- bring their Year 5 class and the required number of accompanying adults to a 75-minute workshop led by the author/illustrator
- bring finished examples of children's own writing to the last half-day training session
- allow the research team access to collect data (for example through observations and interviews);
- allow the evaluation team to visit your school to administer a writing assessment at the end of the project;
- complete a survey for each child in the project class at the beginning and end of the study.

If you are a 'comparison' school, you would need to commit to allowing the project teacher to:

- provide student profile and attainment data and UPNs;

- allow the evaluation team to visit your school to administer a writing assessment at the end of the project;
- complete a survey for each child in the project class at the beginning and end of the study.

Finances

- Schools in the **intervention** group will be asked for a payment of £200 for delivery of the intervention; this is a significantly reduced cost with the rest funded by the EEF as part of the trial;
- Schools in the **comparison** group will receive a payment of £500 at the end of the trial once all commitments above have been completed.

If you would like to know more, or if you have any questions, please contact Charlotte Hacking at CLPE by email at pop@clpe.org.uk or by phone on (020)74013382.

Parent information sheet

POWER OF PICTURES

What is this about?

The Centre for Literacy in Primary Education is currently working on a project funded by Education Endowment Foundation and the Royal Society of the Arts, exploring whether and how their Power of Pictures programme helps to improve outcomes in writing for the children in the schools where they work. The effectiveness of the project will be researched by a team from UCL Institute of Education and the Behavioural Insights Team (the 'evaluation team'). This research has been reviewed and approved by the research ethics committee of UCL Institute of Education.

What will the project look like?

The project is investigating the effect of the Power of Pictures programme, run by the Centre for Literacy in Primary Education (the 'project team'), which will work in schools to deliver their programme, which helps primary school teachers to develop their understanding of the craft of picturebook creation and illustration as a way of raising children's achievement in literacy.

We plan to work with 120 Year 5 teachers during 2018–19. 'Intervention' schools (see below) will participate in the Power of Pictures one year programme. You can find out more about the Centre for Literacy in Primary Education and the Power of Pictures from **https://www.clpe.org.uk/powerofpictures.**

What are 'intervention' and 'comparison' schools?

An important element of these EEF projects is that schools are randomly chosen either to be in the intervention group, who will receive the training this year, or a comparison group, who contribute to the data required for comparison (and might choose to do the training in future).

Whether your child's teacher will receive the training this year ('intervention' schools) or not ('comparison schools') will be randomly decided by evaluators from UCL Institute of Education and the Behavioural Insights Team to help them understand how effective the training has been. If your child's teacher does not receive the training this year, they will receive a payment that will allow them to take part in Power of Pictures in future years (or for other purposes, if they prefer).

What does this mean for me as a parent?

As part of measuring the success of this training programme, your child will be asked to complete a writing test and survey during their normal classes towards the end of the school year. This will take about half an hour. We are doing this test for the purposes of the research project, to help us understand if the Power of Pictures programme helps children like yours with their writing. We will also obtain your child's UPN (Unique Pupil Number) to allow longer term understanding of whether this writing programme worked. We will then save this information in a data format that will prevent anyone from identifying your child.

This data will then by linked with the National Pupil Database (held by the Department for Education, part of the UK Government) and shared with the project team, the Department for Education, the Education Endowment Foundation (EEF, who are funding this research), EEF's data contractor FFT Education and kept in an anonymised form in the UK Data Archive. No information that can identify individual children will be made available to anyone outside these teams and your child's school. This data will be kept securely under password protection. We will not use your child's name or the name of the school in any report arising from the research, and no information that could otherwise identify your child will be made public.

Although we think this project will benefit your child and that they will enjoy being part of it, you have the right to ask us not to use your child's data in this way. Please see the letter that came with this information sheet or contact your child's class teacher. If you have any concerns and would like to know more, or if you have any questions, please contact Louise Jones at the Behavioural Insights Team by email at louise.jones@bi.team or by telephone on 07804494899.

Parent objection letter

Dear Parent / Carer,

Your child's school has applied to take part in research that aims to improve their self-confidence and performance in writing. The Power of Pictures programme, run by a team at the Centre for Literacy in Primary Education (the 'project team'), helps primary school teachers to develop their understanding of the craft of picturebook creation and illustration as a way of raising children's achievement in literacy.

This process will then be researched by a team from UCL Institute of Education, the Behavioural Insights Team, along with some evaluation by the Royal Society of Arts (the 'evaluation team). There's more information on the information sheet that came with this letter.

Not all schools in the study will necessarily take part in Power of Pictures this year. Whether your child's school will take part this year will be decided by the evaluation team at random to help them understand how effective the project has been.

As part of measuring the success of this training programme, all Year 5 children will complete a classroom based writing test and survey towards the end of the year. This will take about half an hour. We are doing this test is for the purposes of the research project, to help us understand if the Power of Pictures programme helps children like yours with their writing.

Your child's name and other data held by the school, alongside their writing test scores, will be collected by the evaluation and project teams. No information that can identify individual children will be made available to anyone outside these teams and your child's school. This data will be kept securely under password protection. We will not

use your child's name or the name of the school in any report arising from the research, and no information that could otherwise identify your child will be made public.

We will also obtain your child's UPN (Unique Pupil Number) to allow us to link up our data with the National Pupil Database (held by the Department for Education) and other official records to understand whether being part of this project is linked with test scores when they are older. This involves us sharing data with the Department for Education (part of the UK Government), the Education Endowment Foundation (EEF, who funded the trial), EEF's data contractor FFT Education and in a form that will prevent anyone from identifying your child to the UK Data Archive.

This research has been reviewed and approved by the research ethics committee of UCL Institute of Education.

If you have any questions you would like to ask before replying, please contact the research team on Louise Jones at the Behavioural Insights Team by email at louise.jones@bi.team or by phone on 07804494899.

Because we are doing this research to improve understanding about what works in improving pupils' education, if you are happy for information about your child to be used in the Power of Pictures research project you do not need to do anything. Thank you for your help with this research, your support is much appreciated.

If you **DO NOT** want your information about your child to be used to understand whether the Power of Pictures programme can help children to write better, please complete the enclosed form and return it to your child's school by Monday 16th April 2018. If you do this then no information about your child will be shared with the evaluation or project teams at any point during the project.

Power of Pictures research programme

(If you are happy for your child to participate in the research on whether this programme improves writing, you DO NOT need to return this form.)

I **DO NOT** wish my data about my child to be collected as part of this research.

Child's name: ……Date of birth: ……

Child's class Teacher: ……..

School:……

Parent name (BLOCK CAPITALS) ……

Parent signature: ……

Date ……

(Please return the completed form to your child's class teacher.)

Revised GDPR privacy information

EEF/RSA EVALUATION OF POWER OF PICTURES

Data Privacy Notice

Dear Parent,

We've previously been in touch because your school is taking part in a project funded by the Education Endowment Foundation (EEF) to understand the potential benefits of Power of Pictures. As part of that information, you were given the opportunity to tell your school not to pass any data about your child to us to be used as part of this project. **Please rest assured that if you contacted the school to make this request, nothing in this letter changes that.**

We wanted to get in touch again to provide you with further details about the way will be handling pupils' data as part of this project. **It is very important to us that that we do this responsibly and providing these details are an important part of that.** They are also important in fulfilling our responsibilities under the UK's data protection laws, which we take very seriously. These require us to provide you with some specific information about **our plans** and **your rights**.

Some of this description involves rather technical terms, which we've left in so you know the official concepts we are talking about. We've tried to keep the explanations as simple as possible. If we haven't managed that well enough and you have any questions now, or at any point during this project, then you should contact a member of the team with the first point of call being Louise Jones at BIT (louise.jones@bi.team) and Jake Anders at UCL (jake.anders@ucl.ac.uk).

Our Plans

- Using pupils' data as part of research is not something we do without thinking about it. Under data protection law, we require a 'lawful basis' for the data processing that we carry out. UCL will be using the lawful basis known as the 'public task' basis, while the lawful basis BIT are using is known as the 'legitimate interests' basis (it is different at UCL and BIT because UCL is a university). To use the 'legitimate interests' basis, we must consider why this is a legitimate interest and inform you of this. Here, it is because our work is for the purpose of promoting the education or well-being of children in England and couldn't be achieved without analysing these test scores. We balance that against your and child's rights by providing you with the right to object to our use of your child's data in this way.
- Your child's name and other data held by the school, alongside the test scores we will collect, will be collected and processed by us for the purpose of understanding how participating in this project has affected their learning. No information that can identify individual children will be made available to anyone outside these teams and your child's school (with an important exception, as explained below). We will also obtain your child's UPN (Unique Pupil Number) from the school to allow us to link up our data with the National Pupil Database (held by the Department for Education) and other official records to understand whether being part of this project is linked with test scores when they are older. This involves us sharing data with the Department for Education (part of the UK Government), the Education Endowment Foundation (who funded the trial), EEF's data processor FFT Education and (in a form that will prevent anyone from identifying your child) to the UK Data Archive.
- The data we hold will be kept securely at all times, transferred using secure (encrypted) methods, and kept on secure computer systems at UCL and BIT's offices under password protection. We will never use your child's name or the name of the school in any report arising from the research, and no information that could otherwise identify your child will be made public.
- Pupils' personal data will be processed by us only for the purposes of this research project. Once that is complete then the data will be securely destroyed from our computer systems. Personal data will certainly not be more than 10 years, in line with UCL's policy on storing research data.

Your rights

- Under data protection law, you have a right to be informed about our plans. This letter, as well as the information that you previously received from your school, are all part of this. You also need to know exactly who is involved in the data processing. In legal terms, University College London (UCL) and the Behavioural Insights Team (BIT) are considered joint data controllers for this project. The law requires our organisations to have named Data Protection Officers, who are ultimately responsible for overseeing data processing that goes on in their respective organisations. UCL's Data Protection Officer is Lee Shailer, who can be contacted via data-protection@ucl.ac.uk. BIT's Data Protection Officer is Lizetta Lyster, who can be contacted via

lizetta.lyster@bi.team. You should contact these individuals if you have any complaints about how we are processing data. However, if you remain unsatisfied, you may wish to contact the Information Commissioner's Office (ICO). Contact details, and details of data subject rights, are available on the ICO website at: **https://ico.org.uk/for-organisations/data-protection-reform/overview-of-the-gdpr/individuals-rights/**

- As noted above, we provided you with the right to object to data processing before schools handed any information over to us. You can also contact us at any point during the project to request information we hold about your child, to request rectification of any information that is incorrect, to stop using their data as part of the project or to destroy their data. If you wish to make such a request or ask any questions about it then please contact us. The best place to start is to contact Louise Jones at BIT (louise.jones@bi.team) and Jake Anders at UCL (jake.anders@ucl.ac.uk).

Once again, we are extremely grateful to you for supporting this project. We hope to learn a lot about the role cultural and arts education can play in supporting pupils' learning.

Yours sincerely,

Appendix Q: Amalgamated logic model for all five LAC trials

This document is available for download at https://educationendowmentfoundation.org.uk

Education
Endowment
Foundation

The Education Endowment Foundation
5th Floor, Millbank Tower
21–24 Millbank
London
SW1P 4QP

https://educationendowmentfoundation.org.uk
@EducEndowFoundn
Facebook.com/EducEndowFoundn