

# Power of Pictures

UCL Institute of Education and BIT

Nikki Shure, Jake Anders, Dominic Wyse,

Florentyna Farghly, Jessica Heal, Pantelis Solomon, Michael Sanders, Daniel Carr



## Evaluation Summary

Age range	9-11
Number of pupils	Approx. 3,600
Number of schools	120
Design	Two-arm school-level clustered randomised trial
Primary Outcome	Writing Assessment Measure (WAM) with double-weighted "Ideas" sub-scale
Protocol date	18 May 2018
Version	1

## Intervention

The Power of Pictures (PoP) is a programme that uses the creation of picture books to develop primary school pupils' writing skills. Through its programme, PoP aims to enhance teachers' understanding of the power of picture books and increase teachers' comfort in and ability to teach using picture books to a range of ages. In the last year, the PoP team has worked with 65 teachers in 40 schools.

The PoP evaluation is part of a broader programme of work entitled 'Learning About Culture', which aims to improve the evidence base around arts-based education programmes. This is coordinated by the Education Endowment Foundation and the Royal Society for the Arts.<sup>1</sup> It 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.<sup>2</sup>

PoP operates as follows. The PoP team selects picture book authors with whom they wish to work. This selection is based on the quality and relevance of a specific book from the author's work, which will be used for the intervention. The PoP team work with the author and the 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 then 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 to PoP. Typically, two teachers from a school are selected to lead on the intervention: one from Year 1 and one from Year 4 or 5.

<sup>1</sup> <https://www.thersa.org/globalassets/pdfs/reports/rsa-learning-about-culture-report.pdf>

<sup>2</sup> For an overarching flow diagram of the programme similarities, please see appendix 1.

The teachers selected from each school receive 2.5 days of training 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 schools attend 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 focussed on character design and development of narrative.

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 back to their class for another six-week period and use the teaching sequences in more detail to focus on the innovation of pupil's own picturebook stories, inspired by the workshop. This is again documented via photos sent to the PoP team.

The third and final half-day of training involves not only the teachers who have previously participated, but also the school's senior leadership team and the literacy coordinator. On this day, reflection takes place as to the effectiveness of approaches and an action plan is formulated for how to continue incorporating picture books and illustration into schools' curriculum, based on successes shared.

### **Significance**

PoP developed out of the Power of Reading research (part of the Centre for Literacy in Primary Education, CLPE) and a broad literature, which established the significance of picture books in the development of literacy (Doonan, 1993; Kiefer, 1995; Nikolajeva and Scott, 2006). This was shown to be especially true for pupils with weak literacy skills. 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 PoP programme has been evaluated with a report released in 2017. This evaluation was highly favourable of the PoP, finding that the "courses [were] a very impressive success" (Horner and White, 2017). The evaluators found the programme to have a lasting effect on pupils and as well as the author-illustrators who participated.

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 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 picture books they realised how much of the story—including information about character and structure—was conveyed in the pictures.
- Picture books 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 and White, 2017).

This EEF evaluation will extend the evaluation already carried out by robustly assessing the impact of participation in the PoP on literacy and writing outcomes under RCT conditions.

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 cultural learning strategies in English schools. The programmes will be supported by Arts Council England.

## Methods

### Research questions

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.

In addition, the evaluation will seek to answer the following questions:

- What is the effect of participating in the PoP over the course of one school year on pupils' writing self-efficacy?
- 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?

In addition, we will look at the long-term effect of participating in the PoP after one further year, looking at results from the end of Key Stage 2 SATS test in English grammar, punctuation and spelling. These long-term results will not be included in the initial EEF report due to the time lag in these assessments, but this analysis will allow us to look at the longer-term effects of participating in the PoP. The results from the long-term outcomes will be reported in early 2021.

### Design

This trial has been designed as a two-armed clustered randomised trial with randomisation occurring at the level of the school. This level of randomisation has been 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 Power of Pictures (Treatment)
- Business as usual (Control)

### Randomisation

Blocking will be used to improve cross-arm comparability of schools and also to improve precision of estimates. There will be four blocks, defined on the basis of class composition by English as an Additional Language (EAL) (high vs. low) and class composition by eligibility for Free School Meals (FSM) (high vs. low). This approach will help to ensure that our treatment and control groups are well balanced in terms of these characteristics, which are likely to be correlated with our outcome measures (EEF, 2015b). High and low EAL and FSM in these definitions will be defined as above and below by the sample median to ensure that block sizes are approximately equal (which may not be the case if we used population, rather than sample, characteristics).

Randomisation will be designed to achieve an equal number of schools in each arm (i.e. 60 control and 60 treatment):

- Each school will be assigned a randomly generated number (setting a stable seed for the random number generation);
- Schools will be sorted by block and random number;
- Schools will be assigned to the treatment arm and to the control arm in turn.

Randomisation will be carried out by UCL in Stata and the code used to carry out the process will be recorded and reported in the final report.

## Participants

120 English state-funded schools will be recruited from the following geographical regions: London, NE England (training located in Newcastle) and East Sussex (training located in Brighton) Within each school, one teacher from Year 5 will be selected for participation in the trial; all pupils in this teacher's class will participate in the trial. It is important that the teacher and pupils who will participate are selected prior to randomisation to ensure that this cannot introduce differences between the intervention and control groups; this will be assured by collection of pre-randomisation data (see the following section on sample size calculations).

In order to be considered, schools will have to agree to provide student data in order that it is possible to apply for data from the National Pupil Database, to identify a lead teacher from Year 5, and to cooperate with the project and evaluation teams during the trial (further details of these requirements are outlined in the Memorandum of Understanding with Schools,).

CLPE will advertise the trial on its website and also approach schools via its existing schools network. Where possible it will aim to recruit schools that have larger populations of individuals receiving Free School Meals (FSM) than the national average of 15.3 per cent of pupils aged 5-10 (DfE, 2016).

## Sample size calculations

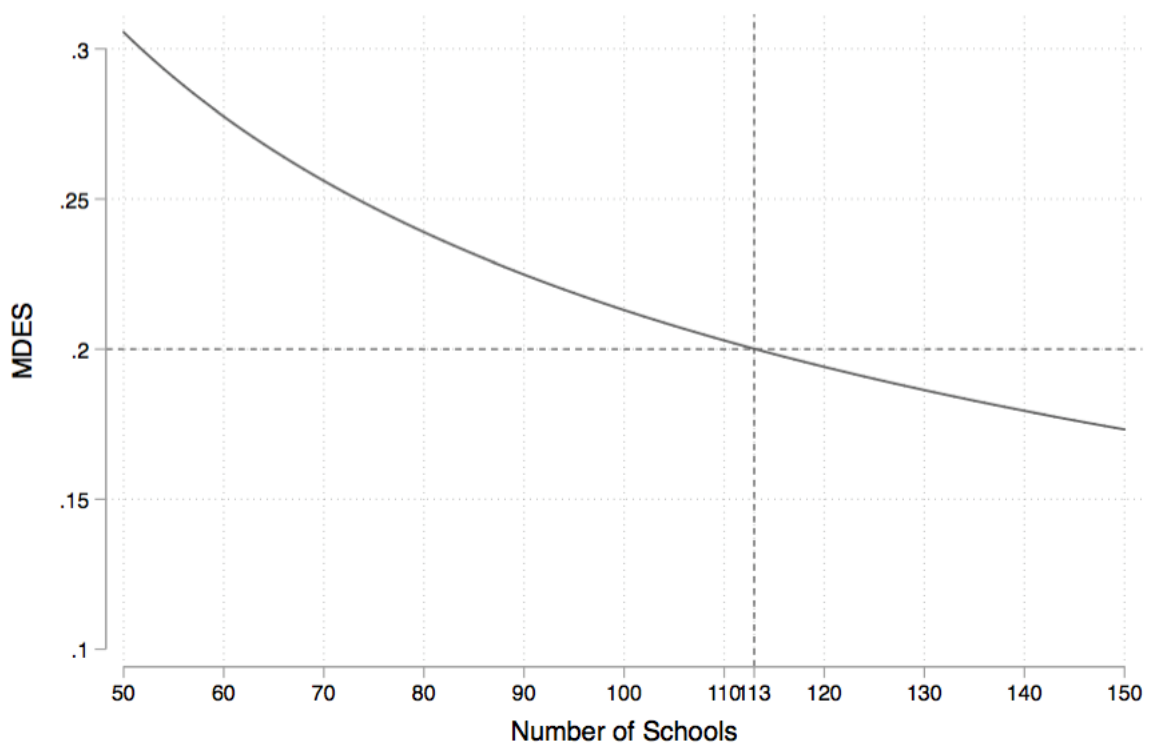
Sample size calculations are based on an estimated Minimum Detectable Effect Size (MDES) of 0.20 and the following assumptions: power of 0.8 for a two-tailed 0.05 significance test, randomisation at school level, an intra-cluster correlation of 0.15<sup>3</sup> (EEF, 2015a) and 25 pupils involved in the trial per school with 10 per cent pupil-level attrition.

An appropriate pre-test/post-test correlation assumption cannot be estimated empirically directly for this trial, since correlation data between the pre- and post-tests used in this trial are unavailable. This is because the pre-test (score in the year 1 phonics screening check; used consistent with EEF policy to use an administrative measure rather than an additional pre-test) has only been in place since 2012, and our post-test (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 and 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 pre- and post-test in this trial is much longer, and we will not be using the WAM itself as a baseline). Our proposed pre-test (score in year 1 phonics screening check) has less variance than would be ideal, due to a degree of bunching between the pass mark (32) and the highest available mark (40). Nevertheless, given its closer temporal proximity to the post test point, we believe it is likely to explain more variance in our post-test than earlier measures available in the NPD (which would have to be measured at the Early Years Foundation Stage). While there is no direct measure of the pre-test/post-test correlation between the WAM and the phonics screening check available, a value has been calculated 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). This value is 0.52. Given the similar time period between pre-test and post-test administration, and the related domain, we believe this estimate is likely to approximate the value that will be observed in this trial. Based on this, we assume that 25% of post-test variance at both pupil- and school-level is explained by the pre-test (equivalent to pre-test/post-test correlation of 0.5).

Figure 1. Minimum detectable effect size estimate as a function of number of schools

---

<sup>3</sup> EEF guidance on ICCs (EEF, 2015a) 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.



These assumptions suggest a requirement of 113 schools to achieve an MDES of 0.2 (see Figure 1). Based on discussions with the PoP team at the set-up meetings, we agreed on a sample size of 120 schools. They confirmed that recruitment of 120 schools and intervention delivery to 60 treatment schools are reasonable and achievable numbers given their capacity. Given the sample size of 120 schools and the assumptions mentioned above this trial should be able to detect an effect of 0.182.

Assuming the FSM sub-group is 15.3 per cent of the total size of the sample (based on data for pupils aged 5-10 from DfE statistics (DfE, 2016) and ignoring that it may be higher if recruited schools are 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 subgroup), there is an estimated minimum detectable effect size for this group of approximately 0.29 standard deviations.

### Outcome Measures

The primary outcome of interest is writing attainment and the secondary outcome of interest is self-efficacy in writing. The “ideation” sub-measure of the same writing self-efficacy measure will be an additional secondary outcome measure.

#### Writing attainment

To measure the primary outcome, we will use 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 measure, relevant within the context of the English educational system. This measure is designed to assess narrative writing in response to a written prompt, to which pupils are given 15 minutes to write. Previous evidence suggests that this measure is reliable (test-retest correlation  $r=0.82$  over 21 days with different prompts) and valid ( $r=0.786$  with Wechsler Objective Reading Dimensions–WOLD–Written Expression subtest) (Dunsmuir et al., 2015). 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: ideas development, organisation and planning, vocabulary, sentence structure and grammar, spelling, punctuation and handwriting. The WAM is unique as an assessment because it incorporates “ideas development”. Given

the nature of the intervention, we will double-weight the score on the “ideas development” dimension. Final scores range from 4 to 32 (after accounting for double-weighting).

### 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 children’s sense of efficacy as a writer. For this reason, we consider writing self-efficacy as our secondary outcome. In addition, self-efficacy has been highlighted in EEF’s review of non-cognitive skills: the evidence “indicate that self-efficacy for a particular task is malleable and that improved self-efficacy is associated with greater persistence, interest and performance” (Schoon & Gutmann, 2017, p.11) and that “the best predictors of specific academic performance and self-efficacy beliefs regarding those specific academic domains” (Pajares, 1996). To measure writing self-efficacy, we will use a version of the measure proposed by Bruning et al. (2013), which has been adapted for primary school pupils with some simplification of language. This involves sixteen 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 plan to use a slightly simplified version of this to better suit the primary school context; in addition, we will request responses on a 5 point Likert scale. Bruning et al. (2013) develop a multi-factor model of writing self-efficacy, however since we do not have specific factors (beyond those listed below) that we hypothesise our intervention to affect, we will use 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.

### Ideation

In addition, we will report differences in the “ideation” sub-measure of the writing self-efficacy measure as an additional secondary outcome measure. This measure was jointly chosen with the RSA and allows us to address secondary research question 2 on “idea generation”. This uses the first five questions of the writing self-efficacy measure and, as such, possible scores range from 5-25 for each child.

### Collection and marking

Both the WAM and the writing self-efficacy measures will be invigilated and collected in summer 2019 by a team of research assistants (RAs) coordinated by the Behavioural Insights Team (BIT); they will also mark the writing self-efficacy measure. RAs will be blind to trial arm assignment of schools. An additional group of RAs, also blind to trial arm assignment of schools, will mark the writing exercises against the WAM scoring sheet. A sub-sample will be independently double marked to allow us to assess inter-rater reliability of the measure.

In addition, we will look at pupil performance on national KS2 SATS tests in English grammar, punctuation and spelling. These results will not be available until 2020, which is after the trial concludes; therefore, this analysis will not be included in the initial report. The results from this outcome are planned to be included in a separate report reflecting on all the projects from this round of funding to be published in early 2021.

### ***Analysis plan***

We will estimate the effect of the trial using a linear model on pupil-level data with school-level clustered standard errors including a school-level treatment indicator, blocking dummy variables (EAL composition and FSM composition) and pre-test covariate (Kahan and Morris, 2012). Scores in outcomes as described in the outcome measures section above will be used in all models.

The coefficient on this treatment indicator will recover the Intention to Treat (ITT) estimate of impact. We will calculate Hedge’s g effect size by dividing this coefficient by an estimate of the pooled total variance of the outcome variable and applying the appropriate correction

factor. 95% confidence intervals will be estimated by inputting the upper and lower confidence limits of the coefficient from the regression model into the effect size formula.

An estimate of the intra-cluster correlations of the outcome measure will be extracted by estimating a variance components model for this purpose.

As noted above, the regression model will include a pre-test variable in order to improve the precision of the estimates. This will vary depending upon the outcome being estimated:

- We will use pupils' marks in the Year 1 phonics check (extracted from NPD PHONICS\_PHONICS\_MARK) as a pre-test for writing attainment outcomes and the KS2 English grammar and punctuation outcome measure.
- We will use assessment of pupils' Personal, Social and Emotional Development skills from the EY Foundation Stage Profile (aggregated scores from NPD FSP\_PSE\_G06, FSP\_PSE\_G07 and FSP\_PSE\_G08) as a pre-test for writing self-efficacy outcomes (including the ideation sub-scale).

Following EEF guidance, we will first test for an interaction of the treatment and FSMever status. If a significant interaction is found, we will estimate a separate model on the restricted sample of only FSMever pupils. This procedure will be carried out for both our primary and our secondary outcomes.

We will estimate treatment effects for compliers using a Complier Average Causal Effect (CACE) analysis using a school-level measure of compliance with the intervention, discussed below.

### **Definition of fidelity/on-treatment minimum**

We outline below the fidelity measure and on-treatment minimum for PoP below. This measure assesses the minimum standards required in order for the delivery team to be satisfied that it is on-treatment – it is not an assessment of quality of engagement. The purpose of this measure is to be able to exclude schools which have not engaged in the intervention in the way we expected, which also provides useful contextual information for the process evaluation. For example, it may help us decide which schools to sample for the case studies.

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

These metrics assess the minimum standards required for the delivery team to be satisfied that it is on-treatment – it is not an assessment of quality of engagement. The purpose of these measures is both to understand the dosage of the intervention, as well as to contextualise the process evaluation. All measures will also double up as continuous measures to assess the range of basic engagement within the sample<sup>[FT4]</sup>, and will help us understand which schools to ask to participate in the case studies as well as provide data to cross reference against the survey results.

## ***Implementation and process evaluation methods***

### **Introduction**

A robust and in-depth implementation and process evaluation (IPE) is vital to ensure we understand the extent to which PoP achieves positive outcomes for young people.

In the first section, we outline the overarching implementation questions that will be explored across all Learning about Culture projects, including PoP. These cross-project similarities in delivery and in what they are aiming to achieve are outlined in the appendix. We highlight, for each question, which dimension or factor affecting implementation it relates to, as specified in the guidance set out by the EEF.<sup>4</sup>

The second section outlines the IPE questions that are unique to PoP.

A flexible research approach will be employed to capture the unifying and distinct elements of the five programmes. We will use similar methods to capture both the overarching IPE questions, as well as the project specific questions.

### **Cultural Learning IPE Questions**

1. In what ways was the programme implemented? What are the barriers and facilitators of delivery (Fidelity)? In particular:
  - a. Senior Leadership Team buy-in;
  - b. Delivery of training – a) the extent to which is it consistent across sites; and, b) whether it appears to be effective in ensuring that teachers understand the aims and main features of the intervention;
  - c. Delivery of the intervention – a) consistent across sites; b) whether it appears to be effective in supporting children’s attainment c) 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. To what extent is the knowledge of arts practitioners delivering the intervention integrated with the pedagogic knowledge of teachers involved? (Implementer support system)
5. How does a pupil’s home learning environment and exposure to books affect their engagement? (Characteristics)

### **Power of Pictures**

Beyond the overarching questions listed above, additional areas which will be important to explore are as follows:

1. 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?
2. In what ways do the illustrators’ delivery of the PoP training vary? How does this variation impact on teachers delivering PoP?

---

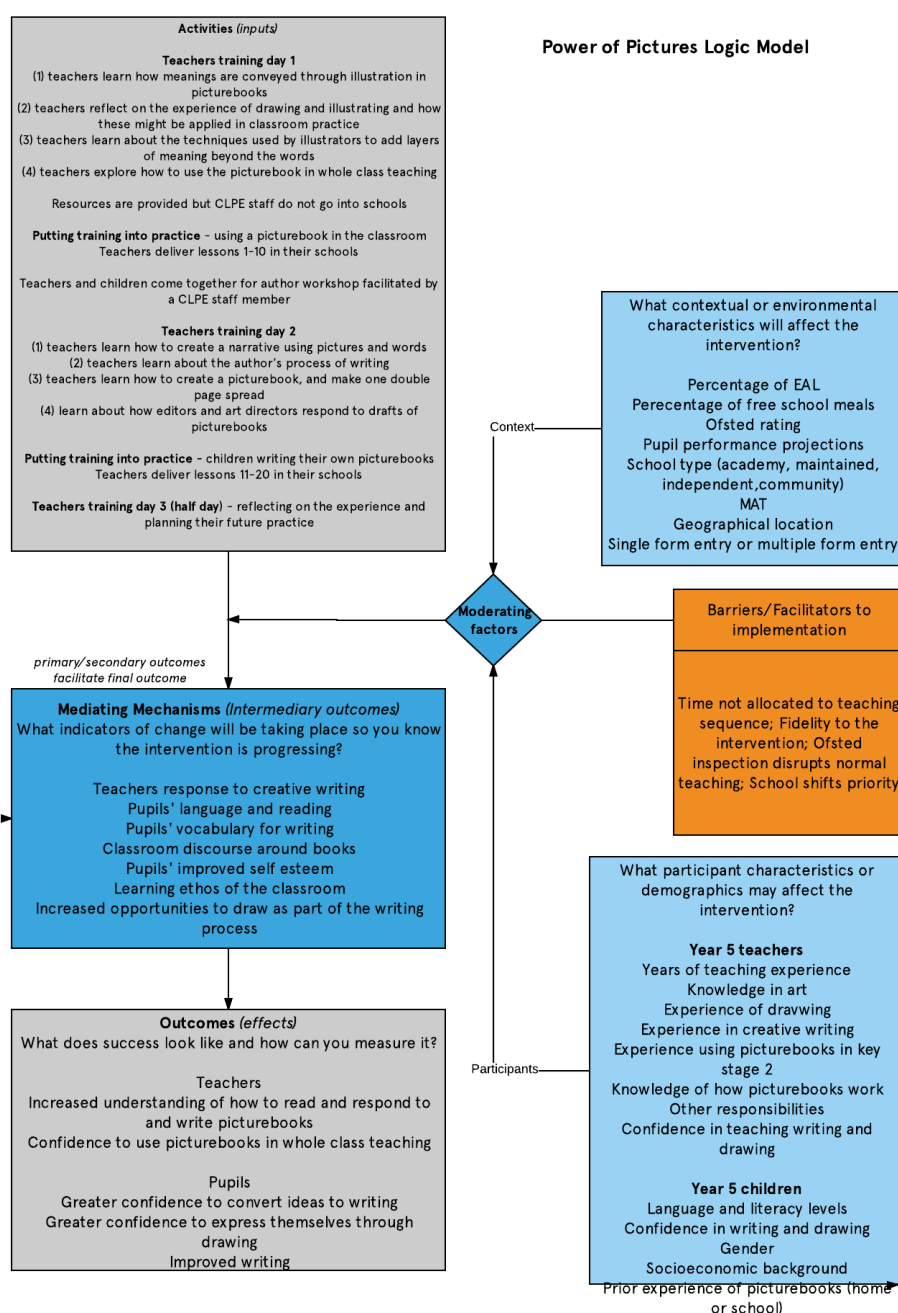
<sup>4</sup> Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R., & Kerr, K. (2016). Implementation and process evaluation (IPE) for interventions in education settings: An introductory handbook. *Education Endowment Foundation (Ed.)*.



3. Does pupils' engagement differ with different author/illustrators? Does this have any relation to their self-esteem and confidence in relation to writing?
4. 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)
5. Are there any unforeseen mechanisms in the intervention that appear to be bringing about change?

## Logic Model

An IDEA workshop was held, utilising the TIDieR framework, to develop a logic model in collaboration with PoP. The Logic Model will be instrumental in directing the IPE. Throughout the IPE, we will attempt to monitor the proposed mediating mechanisms as well as understand the role played by potential moderators. A summary of the similarities across all the logic models for the Cultural Learning interventions can be found in the appendix.



## Methods

A suite of methods will be used to answer the research questions outlined above. These will be analysed in conjunction with the other sources of data to provide an in-depth yet generalisable understanding of the intervention. These methods will be the same across all projects to ensure consistency, but will vary according to the project delivery timetables that are yet to be defined. We will work closely with the PoP team to ensure we conduct the data collection when appropriate.

In addition to main project team input Professor Andrew Burn, specialist in English, Media and Drama; and Professor Gemma Moss, literacy specialist (all at UCL Institute of Education) will be invited to give feedback on the methods.

**Observation of training.** The IPE team will attend and observe at least one author workshop and one training session delivered by the training provider, as well as review the materials used in the courses. Members of our team with expertise and knowledge of arts in education will lead the observations and fieldwork. We anticipate that the project team and/or training providers would also carry out evaluation of the training for their own purposes; where these overlap, and with appropriate consent, we would look to triangulate insights. This will be particularly valuable around measuring engagement in programmes and consistency of training.

**Administrative data.** Working closely with the delivery partners, we will devise measures of engagement in the intervention and triangulate these metrics with the sampling to ensure our case studies (see below) target a variety of intervention settings. These measures may include online metrics, attendance or other relevant engagement related data. This will help us ascertain the feasibility and scalability of projects.

**Case studies of schools.** These will consist of interviews and classroom observations with a subset of approximately 6 schools.

These case studies will consist of

- Teacher interview both before and after the lesson observation
- Observation of a PoP lesson
- Informal interviews with children
- Interview with SLT

The schools will be sampled based on a range of characteristics such as location, Ofsted rating and engagement. Where possible we will interview the Illustrator or delivery staff related to the school. We will use documentary analysis of the resources at the heart of an intervention. Case study is a powerful research strategy to use within sequential explanatory mixed method designs and adds completeness to the exploration of complex issues in situ (Yin, 2013).

**Online surveys.** To gather data from all participating schools, we propose carrying out an online survey of control and treatment schools. The purpose of this survey would be to collect information on “business as usual” schools and classrooms, differences between “business as usual” and intervention classrooms, cost data, and a wider view of implementation and/or impact as measured qualitatively. To encourage participation and

minimise the burden on respondents, it is expected that the survey would take teachers no more than 20 minutes to complete.

### Triangulation

Multiple sources of data will be brought together to best answer the IPE questions. How these methods will be triangulated are outlined in the table below.

<b>Cultural Learning IPE Questions</b>	<b>Methods</b>
<p>In what ways was the programme implemented? What are the barriers and facilitators of delivery (Fidelity)? In particular:</p> <ol style="list-style-type: none"> <li>a. Senior Leadership Team buy-in;</li> <li>b. Delivery of training – a) the extent to which is it consistent across sites; and, b) whether it appears to be effective in ensuring that teachers understand the aims and main features of the intervention;</li> <li>c. Delivery of the intervention – a) consistent across sites; b) whether it appears to be effective in supporting children’s attainment c) whether it appears to facilitate children’s engagement</li> <li>d.</li> </ol>	<p>Survey; Administrative Data; Case studies; Observation</p>
<p>To what extent did the schools engage with the intervention, in line with the intervention aims? (Responsiveness)</p>	<p>Survey; Administrative Data</p>
<p>How was the quality of the intervention perceived by teachers, senior leaders and teaching assistants? (Quality)</p>	<p>Case studies; Survey</p>
<p>To what extent is the knowledge of arts practitioners and other practitioners integrated with the pedagogic knowledge of teachers involved? (Implementer support system)</p>	<p>Case studies; Survey</p>
<b>Power of Pictures Questions</b>	
<p>In what ways are the teachers’ interpretations of the picture books and associated lessons, recommended as part of the CPD programme, reflected in the delivery of the intervention? (Adaptation)</p>	<p>Survey; Case studies; Observation</p>
<p>What appears to be the level of pupils’ interest in different author/illustrators in relation to their motivation for writing?</p>	<p>Survey; Administrative Data</p>
<p>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)</p>	<p>Survey; Case studies</p>

What are the mechanisms in the intervention and to what extent are they bringing about change?	
How does a pupil's home learning environment and exposure to books affect their engagement? (Characteristics)	Survey; Case studies

### **IPE Analysis**

Structurally, this will draw upon the analytical strategy of multi-case studies – whereby a programme is first coded individually and then a large cross-sectional analysis is conducted, which encompasses all programmes (Stake, 2013). This deductive analysis will be conducted on Nvivo by the lead researchers, who will co-code 3 transcripts to ensure coder similarity and robustness of coding framework. Codes will be pre-specified in a coding framework which reflect the research questions, but additional codes will be created as new themes emerge.

The analysis will be conducted in stages, first on the school, or case level, then across the cases involved in the trial. Finally, a cross-project analysis of the Cultural Learning aspects of the data will be conducted to ensure we identify significant patterns relevant to all interventions. This will take the form of a flexible, yet robust, thematic framework, which will include elements that are unique to each, but also relevant to all projects. It will be important to understand how the same theme may be manifested in a different way for different programmes (Bazeley, 2013).

### **IPE Data Collection Timeline**

We understand that each project will follow a similar delivery schedule, with variation in the numbers and timing of training sessions across the year. This similarity allows us to map our data collection activities on to one timeline. We have arranged the timeline by term as the PoP team are yet to specify exact timings for their programme delivery. We can therefore consider this an indicative schedule of events across the academic year of 2018-19.

<b>Date</b>	<b>Item</b>
<b>Autumn Term 2018</b>	Observation of Illustrator training
	Collection of baseline survey to measure school buy-in and teacher attitude towards intervention
	Collection of school characteristics
<b>Spring Term 2019</b>	Observation of mid-point training

	Collection of fidelity data to inform case study sampling
	Finalise sampling strategy
	Conduct in-school case studies
<b>Summer Term 2019</b>	Conduct in-school case studies
	Administer end of intervention survey
	Conduct analysis

## Costs

An estimate of the per-pupil cost of the intervention will be calculated by the evaluation team. This estimate will focus on cost from the perspective of a participating school and will be based on the direct, marginal costs of implementing the intervention. This could include anything which the school needed to pay for beyond the business as usual.

The cost estimates will make use of information from the project team (particularly with regarding the actual cost of delivering the intervention, e.g. the cost of providing the training), as well as that collected directly by the evaluation team from schools about the costs of preparing and implementing the intervention. Information on costs, especially any hidden costs or resource implications, will be explored through the process evaluation as part of the interviews with teachers and school visits. The purpose of collecting such data in the process evaluation would be to identify the main areas of expenditure required by the project. This process will also help to establish whether it may be appropriate to include any questions on costs/resource use in the survey. This will need to strike a balance between collecting sufficient cost information and not damaging response rates; it will also need to take account of whether a teacher is well placed to provide accurate information on particular types of costs.

Time spent by schools, such as the amount of time for which schools need to arrange cover for teachers to attend training will be reported separately from the financial costs. Any costs in terms of prerequisites will also be considered, for example pupil travel costs or other resources. Control group schools will also be asked about the time they invested in CPD, to ascertain how much time above and beyond business as usual is needed. We may also triangulate national data on this if available.

An estimate of cost per pupil per year will also be calculated based on the trial period, as once trained, teachers would also be able to deliver the programme in subsequent years. Any costs associated purely with the trial will be excluded.

## Ethics and registration

Ethical approval has been sought following UCL Institute of Education staff ethics approval procedure. It was approved on 23 January 2018.

Personal data for this trial will be processed under the public task provision of the GDPR. Nevertheless, parents will be provided with the option to object to this processing of their child's data, which we will respect. This use of data has been allocated the following UCL Data Protection Registration Number: Z6364106/2017/12/54 social research.

This trial protocol has been pre-registered at [www.controlled-trials.com](http://www.controlled-trials.com), and assigned an [International Standard Randomised Controlled Trial Number \(ISRCTN\)](#) of XXX.

## Personnel

Project team

Charlotte Hacking, Farrah Serroukh (CLPE)

Evaluation team

Jake Anders, Nikki Shure, Dominic Wyse (UCL), Florentya Farghly, Jessica Heal, Michael Sanders (BIT)

The teams will have the following roles within the evaluation:

Design of the trial

- Sample size calculation – UCL
- Refinement of randomisation approach – UCL

Delivery of the intervention

- Recruitment of schools – CLPE
- Delivery of intervention – CLPE

Measurement of outcomes

- Writing outcomes - BIT
- Writing self-efficacy outcomes – BIT
- NPD application and linkage – UCL

Impact analysis – UCL (lead) and BIT

Qualitative analysis – UCL (lead) and BIT

## Risks

The data security policies of UCL and BIT and the Data Sharing Agreement between BIT and UCL are included with this protocol.

Some of the key risks are summarised in the table below:

Issue/risk	Risk level	Action to address issue/reduce risk
Dropout / non-compliance of settings	Medium	We want to avoid attrition of schools from the project as much as possible. We plan to minimise attrition by ensuring that schools that sign up are committed (by asking them to sign a Memorandum of Understanding). Keeping them informed of progress and providing reminders of next steps will be important for retention. The project team as a matter of course monitor changes in key personnel to ensure ongoing commitment. Minimising the data collection burden on schools will also be important for retention. We will also randomise

		only after schools have followed consent collection procedures, providing the necessary student data.
Difficulty in collecting data needed prior to randomisation (i.e. pupil data and consent)	Medium	Data will be submitted directly to BIT who will conduct quality checks prior to its acceptance and a school being considered eligible for randomisation. This will also be subject to quality assurance by the UCL team, including random checks of a sample of data collection spreadsheets during this process ensuring no missing data or discussing where this has been unavoidable and understanding the reasons for this.
Difficulty recruiting schools	Medium to high	We are confident that the project team will convey the importance of the evaluation to settings and the value to them of taking part. To understand whether recruited settings are atypical in some way (which would affect external validity), we ask that the project team keep records of settings approached and, where possible, of reasons for not participating.
Withheld consent to link to NPD	Medium	<p>We plan to collect the necessary data to allow this long-term follow up. We believe this processing of personal data is justified under the legitimate interests/public purpose of data protection regulations. Nevertheless, we will offer parents the opportunity to opt their child out of all processing of their data. We believe we should be able to provide the necessary information to parents in treatment and control settings and do not anticipate high or non-randomly varying levels of opt-out.</p> <p>There is some risk that regulatory change (introduction of the GDPR) may change DfE's attitude towards allowing access to NPD data on this basis. We think there are minimal steps we can take to mitigate this risk directly (short of changing to unambiguous-opt-in-consent, which has its own drawbacks). However, our primary analysis models only rely on NPD data to improve precision; as such we should still be able to recover unbiased estimates, albeit with lower levels of precision.</p>
Missing Outcome Data	Medium	<p>For directly collected assessments, attrition is a potential risk. BIT and UCL will ensure schools and research assistants understand the need to collect post-test measures for as many students as possible in order to maximise internal and external validity.</p> <p>Schools will also be contacted sufficiently far ahead of data collection window to ensure we arrive at a convenient time for RAs to visit and run the writing assessments.</p> <p>RAs will report to the BIT project coordinator the number of children not able to sit the assessments after each visit. If the rate is high (&gt;5% of sample) the project coordinator will contact the school for further detail if required and alert UCL, the EEF and project team.</p> <p>Linking to children's outcomes in the NPD for long-term follow-up offers some protection against attrition but not for the primary outcomes of the project. This is not entirely without risk as it relies on the ability to identify children in the NPD (see above).</p>
Parent and teacher concern about	Medium	Communications to schools (during recruitment) and parents (when obtaining consent) will emphasise that these assessments will be

'over-testing' in KS1 and reception years.		fairly short and their children will not be judged upon the outcomes. As such, they will be kept low-stakes and low-pressure for pupils.
Problematic randomisation	Medium	When randomising clusters rather than individuals, the chances of a 'bad draw' increase because of the reduction in the number of units being randomised for a given number of participating schools. To protect against this, we plan randomisation within blocks, as described in the Randomisation section.
Treatment variation	Medium	We view this not so much as a risk but as the reality of implementing such an intervention. The impact estimates (Intention to Treat) therefore relate more to the type of treatment likely to prevail in practice rather than the type of impact that could be seen were it possible to achieve laboratory-type conditions. Nevertheless, understanding treatment variation is important and will be explored through CACE analysis of the on-treatment sample as well as being a key focus of the implementation and process evaluation.
Unexpected absence or loss of team members	Low	The team will substitute for each other during any short-term absence. In the event of longer periods of unplanned absence or departure, we will recruit replacements. Both BIT and UCL have other experts in evaluation and education who could substitute for members of the team, should this be necessary.

## Timeline

Date	Activity
November 2017-February 2018	<b>Recruitment and preliminary data collection (Project Team)</b> The PoP team will begin recruitment halfway through the Autumn term 2017 using its existing network of schools and continue into early 2018.
October 2017 – February 2018	<b>Pre-Randomisation Data Collection (BIT and Project Team)</b>
May 2018	<b>Randomisation (UCL)</b> Evaluation team will randomise schools and inform PoP of outcome.
October 2018	<b>Intervention begins</b> The intervention will begin in September 2018 and will be delivered by the PoP team and its author-illustrators.
October 2018	<b>IPE begins</b> UCL will lead on process evaluation concurrently with the intervention.
May- July 2019	<b>Data Collection</b> Pupils' writing and self-efficacy outcomes will be measured by BIT. These assessments will be marked by PCGE students at UCL in a process overseen by BIT.
June 2019	<b>Intervention ends</b> The PoP team will conclude delivery of the intervention at the end of the school year.
August-October 2019	<b>Data analysis</b> UCL will lead on the data analysis following the data analysis plan outlined in this trial protocol.



<b>October-December 2019</b>	<b>Report writing</b> UCL and BIT will collaborate on writing up the data analysis and the IPE into a final report to be submitted to the EEF by January 2020.
<b>October 2020</b>	<b>KS2 Outcomes Available for Follow-Up Analysis (UCL)</b>

## References

- Bruning, R., Dempsey, M., Kauffman, D., McKim, C. & Zumbrunn, S. (2013) Examining Dimensions of Self-Efficacy for Writing. *Journal of Educational Psychology*, 105(1), 25-38.
- Catterall, James S. (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 Adults*. Los Angeles/London: Imagination Group/I-Group Books.
- Catterall, J. (2012) *The Arts and Achievement in At-Risk Youth: Findings from Four Longitudinal Studies*. Research Report 55, National Endowment for the Arts.
- Department for Education (2016). *Schools, pupils and their characteristics: January 2016*, SFR20/2016. London: England.
- Department for Education (2017). *Progress in International Reading Literacy Study (PIRLS): National Report for England*. December 2017.
- Doonan, J. (1993). *Looking at pictures in picture books*. 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(2015) 1-18.
- EEF (2015a). Intra-cluster correlation coefficients. Retrieved from [https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing\\_a\\_Protocol/ICC\\_2015.pdf](https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Protocol/ICC_2015.pdf) on 9/1/2017.
- EEF (2015b). Policy on analysis for EEF evaluations. Retrieved from [https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing\\_a\\_Research\\_Report/2015\\_Analysis\\_for\\_EEF\\_evaluations.pdf](https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Research_Report/2015_Analysis_for_EEF_evaluations.pdf) on 19/1/2017.
- Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, 99(3), 445-476.
- Horner, S. & White, J. (2017). *Power of Pictures: Final Project Evaluation*. Centre for Literacy in Primary Education (CLPE).
- Kahan, B. C., & Morris, T. P. (2012). Reporting and analysis of trials using stratified randomisation in leading medical journals: review and reanalysis. *BMJ*, 345, e5840.
- Kiefer, B. Z. (1995). *The potential of picturebooks: From visual literacy to aesthetic understanding*. Merrill.
- Murphy, V., Kyriacou, M. & Menon, P. (2013) *Profiling Writing Challenges in children with English as an Additional Language (EAL)*. Department of Education, University of Oxford.
- Nikolajeva, M., & Scott, C. (2013). *How picturebooks work*. Routledge.
- Schoon, I. & Gutman, L. (2013) *The impact of non-cognitive skills on outcomes for young people: Literature review*. UCL Institute of Education/Education Endowment Foundation.

Warwick Commission (2015). Enriching Britain: Culture, Creativity and Growth. The 2015 Report by the Warwick Commission on the Future of Cultural Value. University of Warwick.

## **Appendix 1: Overarching IPE similarities**

### **Similarities across projects**

The logic models from the 5 cultural evaluations were compared to understand their similarities and differences. From this, an amalgamated flow chart was designed to show the general route that all the programmes can take (Figure 2).

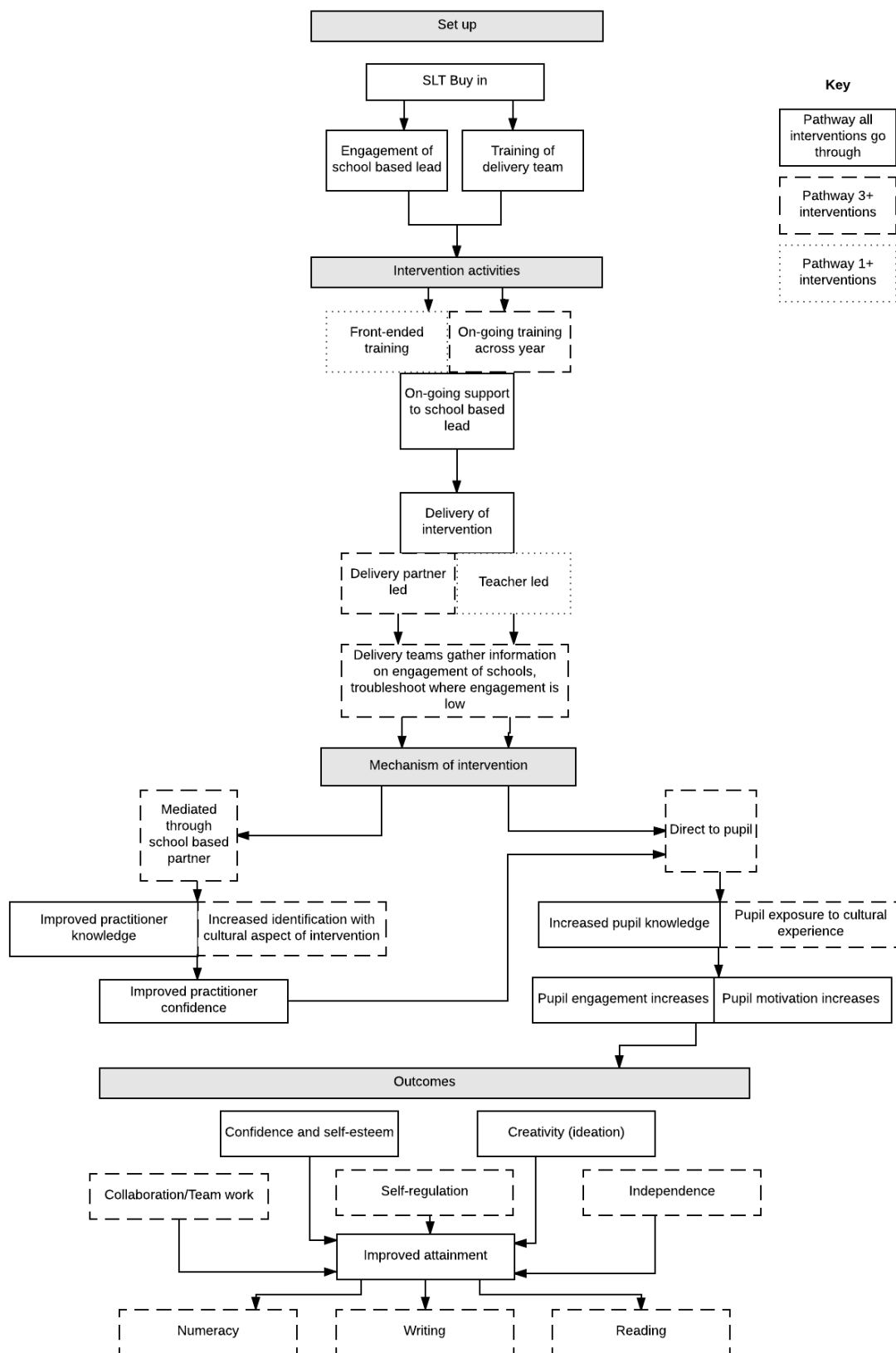


Figure 2 Amalgamated logic model of the 5 interventions

From Figure 2, we can see that the following are standard across all 5 interventions:

**Implementation Similarities**

1. Senior leadership buy-in
2. On-going (yet varied) support from delivery team staff - relationship with school, and teachers or teaching assistants.
3. Training days for teachers or teaching assistants
4. Delivery teams gather information which helps them understand how the schools are engaging in the intervention - to what extent it can we use this to gather fidelity information?

When considering the differences in implementation there are two possibilities which all of the five interventions take: 1. The intervention is mediated through school-based partners, or 2. The intervention is delivered direct to pupils. These two possibilities should be measured in a standardised fashion as they may have implications for how arts-based programmes are designed in the future. These 'options' are outlined below:

- 1) Training model - front-end loaded and/or on-going across the year
- 2) Direct delivery of intervention - via a member of school staff or via a delivery partner
- 3) Mechanisms of change - mediated through a member of school staff or delivered directly to pupils
- 4) For writing orientated interventions, the extent the practices reflect robust evidence of what works?

### **Moderating factors**

Across the 5 interventions, several common moderators emerged from the logic model IDEA workshops. We will aim to capture these systematically when drawing up the MOU with the schools. Of all the 29 different moderators outlined, we will systematically capture those referenced by 4 or more of the projects. These are as follows:

- 1) School Ofsted rating
- 2) Current activities relevant to the intervention
- 3) Pupil SEND/EAL
- 4) Teacher/TA experience (years)
- 5) Teacher/TA background knowledge in arts-related programmes

### **Mediating factors**

There was generally much less overlap between projects overall in relation to mediating factors, and the 43 mediating mechanisms listed (although many between-project similarities). The only ones which were relevant for 4 or more of the projects were broad, and the first is being captured in some of the projects already. The second, creativity, will also be captured as part of the overarching Ideation measure.

- 1) Improved pupil self-efficacy
- 2) Improved creativity