

Speech Bubbles

Behavioural Insights Team and UCL Institute of Education

Nikki Shure, Jake Anders, Dominic Wyse, Florentyna Farghly, Kimberly Bohling, Pantelis Solomon

PROTOCOL UPDATES:

p. 8 – additional information added on Social Skills subscales

p. 9 – reference added as footnote

p. 17 – ISRCTN number and legal basis for accessing data added

p. 18 – study team updated

p. 20 – updated risk mitigation strategies for NPD access and missing outcome data

Evaluation Summary	
Age range	Year 1 and Year 2 (ages 5 to 8)
Number of pupils	800 to 1040
Number of schools	20 to 26
Design	Individually Randomised Controlled Trial
Primary Outcome	Reading attainment and oral communication
Protocol date	6 February 2019
Version	2

Intervention

The Speech Bubbles intervention aims to improve children's reading, communication and social skills by providing them with weekly creative drama sessions. This is an intervention targeted at pupils with below expected communication and social skills. During the sessions trained practitioners encourage children to tell, act out and reflect on their own stories by creating a safe and playful environment, promoting children's communication, confidence and wellbeing. This is based on the Helicopter Stories pedagogical approach.¹ Speech Bubbles 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.² 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.³

Speech Bubbles currently runs in 44 schools across the country and has been the subject of numerous research projects.^{4,5} The programme was developed by the London Bubble Theatre Company with professional support from speech therapists, educational psychologists, and Southwark Pupil Development Centre.

The model that will be tested in this programme comprises 24 weekly drama sessions over the course of three terms. Weekly 45-minute creative drama sessions will be delivered at schools by a Teaching Assistant (TA), selected by the school, and a trained freelance drama

¹ Lee, T. (2015). *Princesses, Dragons and Helicopter Stories: Storytelling and story acting in the early years*. Routledge.

² <https://www.thersa.org/globalassets/pdfs/reports/rsa-learning-about-culture-report.pdf>

³ For an overarching flow diagram of the programme similarities, please see appendix 2.

⁴ Barnes, J. (2014). Drama to promote social and personal well-being in six-and seven-year-olds with communication difficulties: the Speech Bubbles project. *Perspectives in Public Health*, 134(2), 101-109.

⁵ Price, H. & Ansong, E. (2016). *'Speech Bubbles' Drama Intervention Programme Preliminary Executive Summary of Effectiveness*. University of East London.

practitioner, recruited by London Bubble and trained in the Speech Bubbles approach, to two mixed groups of 10 pupils from Year 1 and Year 2. Each session follows a clear and repeated routine centred around the telling and re-enacting of stories that are told by the children. The sessions include activities that support expressive language, receptive language and turn taking.

Speech Bubbles practitioners use a narrative stimuli pack to support the narrative of the stories, along with emotional faces stimuli to support children who struggle to convey emotion via language. The sessions conclude in story collection for the next week's session. After the session, time is allocated for the practitioner and TA to discuss and plan. There will be two open parent/carer sessions during the programme and one teacher session. The intervention will commence with a CPD certified induction day. This day will be held centrally and will bring together school support staff and the drama practitioners who will be delivering the intervention. Additionally, all participating schools will be offered a short training session for all KS1 staff in their school. This will be delivered in the autumn term as a 45-minute twilight session. A half-day practical evaluation session will occur mid-year for the TA and practitioner to reflect on the progress of the intervention.

In Easter 2018, participating schools will be asked to refer 40 children who will be moving into Year 1 and Year 2 in 2018/19 in accordance with the referral guidance, and to give the parents/carers of these children the opportunity to opt-out of the study in the 2017/18 school year. Referral will be on the basis of speech, behaviour, and language difficulties.⁶ London Bubble will provide ongoing support and supervision to each participating school.

Significance

There is compelling non-experimental evidence to suggest that programmes similar to Speech Bubbles can improve academic attainment across several language-related areas. A large-scale review of evidence of the effect of arts education on academic achievement identified a causal link between classroom drama and an improvement in a variety of verbal areas.⁷ Large effect sizes were found in domains of written understanding and recall of stories, and moderate effects in areas including oral understanding, reading achievement, oral language and writing. Moreover, several studies demonstrated that drama helps to develop verbal skills that transfer to new materials, not just those practiced during drama sessions.⁸

In 2013 the Shine Trust funded an evaluation of the Speech Bubbles programme. At the time, the programme was running in 29 schools in disadvantaged areas of London and North West England.² Multiple sources of data were used in the evaluation including: school records, the reports of theatre practitioners, parents, class teachers, speech therapists and a team of independent teachers were also consulted. Feedback from teachers suggested that over 80% of Speech Bubbles children showed improvements beyond expectation in their learning and their speaking and listening skills, and in over 50% the progress was either clear or striking. In addition, teachers reported that 85% of participating children showed progress in their emotional and conduct behaviour.

Another evaluation, commissioned and funded by the London Bubble Theatre Company and conducted by the University of East London, found that children who participated in the Speech Bubbles programme made very good progress relative to a comparison group. Children in the treatment group showed significantly faster improvements in their speech,

⁶ See Annex 1 for Speech Bubbles referral guidance

⁷ Hetland, L., & Winner, E. (2001). The arts and academic achievement: What the evidence shows. *Arts Education Policy Review*, 102(5), 3-6.

⁸ Podlozny, A. (2000). Strengthening verbal skills through the use of classroom drama: A clear link. *Journal of Aesthetic Education*, 34(3/4), 239-275.

language and communication development as measured using the Communication Trust's Primary Speech, Language and Communication progression tool. In particular, it was found that the intervention had a significant impact on the following skills: understanding spoken language, storytelling and narrative, and social interaction.³

This trial provides an opportunity to build on this evidence and explore the impact of Speech Bubbles using experimental methods that allow for more definitive attribution of any measured change. 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 test whether the Speech Bubbles intervention improves reading attainment for pupils in Years 1 and 2 over the course of one year.

The evaluation will also address the following questions:

- Does the programme impact pupils' social skills?
- Does the programme affect pupils' perception of their ability to generate and use ideas in their work?⁹
- What is the impact of the programme on pupils' narrative oral skills?

Design

This will be an individually randomised controlled trial. The trial aims to recruit at least 800 children across 20 schools, with pupils randomly allocated to either the treatment arm (who will receive the programme) or the control group. Forty children will be recruited within each school and assigned to the treatment and control conditions at a 50:50 ratio. Pupils in the control group will continue on a 'business as usual' basis.

Parents will be able to opt their child out of all data collection and analysis elements of this study. These children will not be included in the National Pupil Database (NPD) request, or any analysis, but will still be eligible to participate in Speech Bubbles, and thus will be included in the randomisation. Once randomised, information on these pupils will be deleted and no end-line assessments for them will take place.

The process by which children enter into the study (or are randomised for those whose parents opt out) will be as follows:

1. Teachers refer 40 pupils across Reception and Year 1, as discussed in the Intervention section.
2. Teachers will distribute opt-out forms to the parents of the 40 children.
3. Once the necessary time has elapsed for opt-out return, teachers will upload a spreadsheet of pupil data to BIT containing:
 - a. First name, last name, date of birth (DOB), Unique Pupil Number (UPN), Free School Meal (FSM) status, teacher name and class ID.
 - b. For the children whose parents have opted them out, the school will only supply their first and last name, alongside teacher name and class ID, in order for us to randomise them to a group, but they will not be part of the data collection. No other details will be requested.

⁹ Addressed in the Impact and Process Evaluation section of the Trial Protocol

4. BIT will randomise the 40 pupils into trial arm conditions as per the section following. A running log will be kept of the number of pupils who've opted out of the study, and updated power calculations will be provided at regular intervals.

Children whose parents opt them out will be excluded from the study and have no further data collected.

Randomisation

Pupils will be randomised into trial arms within schools following the completion of the opt-out process and data transfer. To accommodate schools completing this at different times, randomisation will be performed on a rolling basis by school. Hence, school and year level (Years 1 and 2) will be our blocking variables. The randomisation will proceed in the following steps for each school:

1. If there are more than 40 children referred,¹⁰ a random number will be generated for each and those with the highest 40 random numbers will be selected for participation in the trial. Those who are not selected will not be randomised, or have their data collected or analysed.
2. Children in the school will be stratified by year level.
3. Within each block, a random number will be generated for each child.
4. The children with the highest X numbers in the Year 1 block and the highest Y numbers in the Year 2 block will be assigned to treatment.
 - a. X and Y will sum to 20 and reflect the share of children referred by year level. For instance, if 16 children are referred from Year 1, and 24 from Year 2, X would take a value of 8 and Y a value of 12.

We will progressively check that the resulting control and treatment groups are balanced in terms of the absolute number of children who will not participate in data collection due to their parents opting them out of the study. This is to ensure the number of children allocated to each trial arm does not become unduly unequal.

Randomisation will be conducted by BIT staff using the data analysis and statistical software Stata. The code used to carry out this randomisation will be recorded and reported in the final report.

Participants

In order to participate in the study, schools will need to:

- be located in North West England, South London and East London (for programme delivery purposes);
- be at least a two-form entry school (to reach the required sample size across a smaller number of schools)¹¹;
- have discussed participation with Speech Bubbles and signed a Memorandum of Understanding (MoU) detailing the conditions of participation (opt-out process, pupil data provision, end-line assessment, participation in IPE activities etc.); and
- be able to refer 40 children into the study.

Schools with an average or above average share (14.1%¹²) of FSM children will receive priority in recruitment.

¹⁰ Schools will be discouraged from doing this and asked to prioritise referring those students they believe would most benefit from the intervention.

¹¹ With an exception of one pre-agreed school

¹² https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/650547/SFR28_2017_Main_Text.pdf

Pupil-level eligibility will be determined as follows:

- pupils must be in Years 1 and 2 in the year of intervention delivery (2018/19);
- pupils must be referred into the programme on the basis of Speech Bubbles referral guidance. This guidance is attached in Annex 1; and
- have not been opted-out of the study by their parents.

Sample size calculations

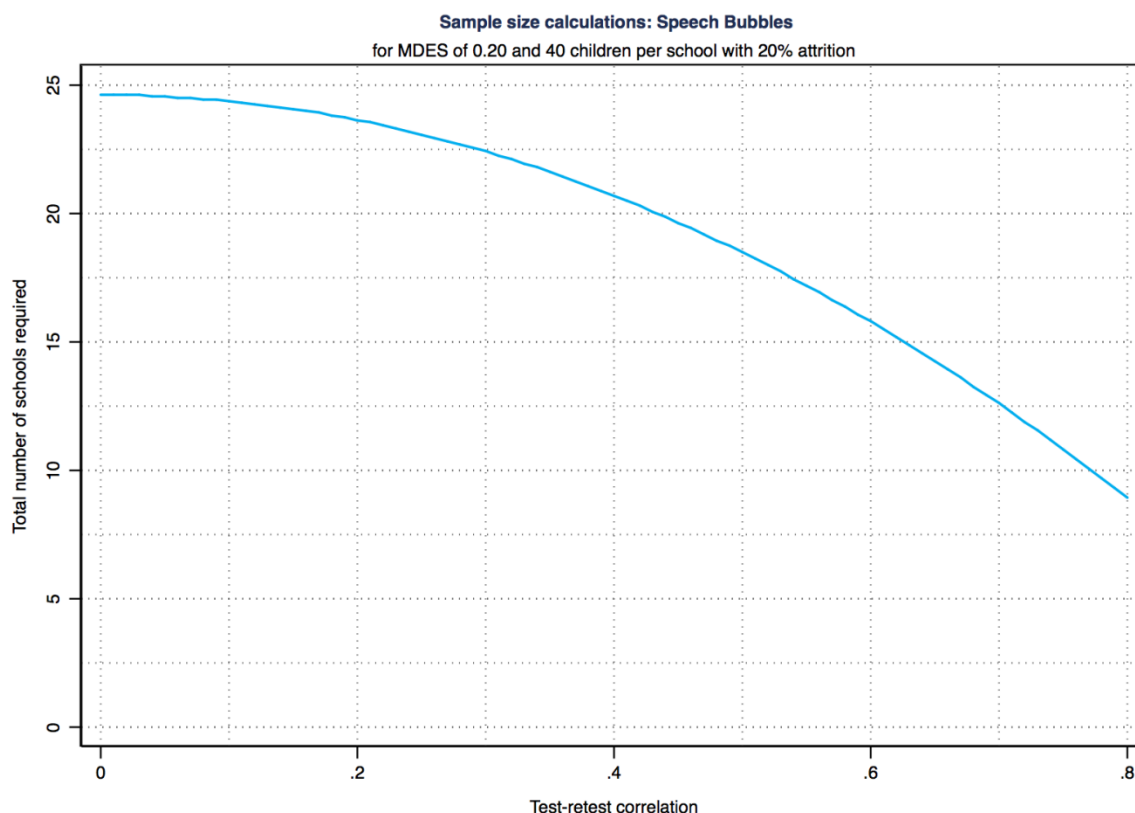
Sample size calculations return the total number of schools necessary to run the trial. They are based on the following assumptions with reference to the primary outcome measures (reading attainment and oral communication):

- **Randomisation will be performed at an individual level.** This means that referred pupils will be randomly allocated to either the treatment or the control group.
- **There will be two trial arms (treatment and control) with 40 children in each school split equally into control and treatment groups.**
- **Of the 40 children randomised in each school, we assume that 20 per cent will have opted out or not participate in the collection of the end-line outcome measures for various reasons (e.g. attrition due to changing school, prolonged absence, inability to engage with the end-line assessments etc.).** This estimate is based on the 15% standard post-randomisation attrition rate in EEF studies, plus an additional allowance for children being opted-out of the study (5%). This reduces the minimum number of children per arm for the purposes of sample size calculations to 17.
- Hypotheses
 - **Null hypothesis:** There is no difference in standardised Progress in Reading Assessment (PIRA)/Renfrew Bus Story scores between children who participate in the Speech Bubbles drama intervention and those who do not.
 - **Alternative hypothesis:** There is a difference in standardised PIRA/Renfrew Bus Story scores between children who participate in the Speech Bubbles drama intervention and those who do not.
- **The required minimum detectable effect size (MDES) is 0.20 standard deviations (Cohen's d).** This specifies the minimum effect size our trial is powered to detect, in terms of a given standardised difference between two means of a continuous outcome measure.
- **Power: 80%; significance level: 5%.** These are standard assumptions in social policy trials.
- **Test-retest correlation.** As we will use Early Years Foundation Stage Profile (EYFSP) scores as a baseline when analysing our primary outcome measures, the predictive power of this baseline will also factor into our sample size calculations. We are able to estimate this value for the reading assessment PIRA using unpublished Fisher Family Trust (FFT) analysis of the test-retest correlation coefficient of EYFSP score and PIRA assessments collected at the end of year 1 for a prior EEF trial (ABRA: Online Reading Support).¹³ This provides a value of 0.61. However, this estimate uses data from just one study, and as our present study is on a specific targeted population, we may find the actual correlation to be significantly lower. For Renfrew Bus Story, assessment of oral communication skills, we do not have information on test-retest correlation. As such, we conduct power calculations for a range of possible values.

Given the uncertainty around test-retest correlation, we present below a graph showing the impact of this variable on trial power. Note that even if a correlation of zero is observed, the study is still adequately powered if the recruitment target of 25 schools is achieved. This

¹³ https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/EEF_Project_Report_ABRA.pdf

recruitment target provides adequate margin for error given the uncertainties raised in our sample size calculations, and shields the study from the risk of under recruitment of schools at the point of randomisation.



Assuming the FSM subgroup is 14.1 per cent of the total sample (based on data from Department for Education statistics¹⁴ and ignoring that it may be higher if recruited schools are in more disadvantaged areas and referred pupils may be more likely to be on FSM), and maintaining all other assumptions (and specifying the expected test-retest correlation coefficient value to be 0.30), the estimated MDES for this group is approximately 0.51 standard deviations.

Outcome Measures

The primary outcome measures are reading attainment and oral communication skills (specifically narrative recall), with secondary measure being social skills.

Reading

To measure reading attainment, we will use the PIRA¹⁵ by Rising Stars. PIRA is a standardised and well-known test of pupil's reading attainment which has been used in two EEF evaluations to date.^{16,17} It measures reading ability in the following areas: phonics, literal comprehension, and reading for meaning.

¹⁴ https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/650547/SFR28_2017_Main_Text.pdf

¹⁵ <https://www.risingstars-uk.com/Series/Rising-Stars-Pira-Tests>

¹⁶ McNally, S. (2016). *Evaluation Protocol: An Evaluation of Teaching Assistant-Based Small Group Support for Literacy*. London, United Kingdom: Education Endowment Foundation. Retrieved from https://v1.educationendowmentfoundation.org.uk/uploads/pdf/Digital_-_Small_Group_Support_for_Literacy.pdf.

¹⁷ McNally, S., Ruiz-Valenzuela, J., & Rolfe, H. (2016). *ABRA: Online Reading Support*. London, United Kingdom: Education Endowment Foundation. Retrieved from https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/EEF_Project_Report_ABRA.pdf

Another strength of the PIRA is that tests are produced at a variety of difficulty levels, graduated by school term (e.g. 'Spring Reception', 'Summer Reception', and 'Autumn Year 1'). As the intervention is targeted at children with speech and language difficulties, we will use a test one stage back from that which would normally be used.

End-line PIRA assessments will be conducted during May - June 2019 by trained research assistants (RAs) who will be blind to trial arm assignment. Rising Stars, the publisher of PIRA, will mark the assessments. Analysis will use raw PIRA total scores (0-25).

Oral communication

During the end-line data collection, pupils' oral narrative skills will be assessed using the Renfrew Bus Story¹⁸ test. The Renfrew Bus Story is a short standardised test that assesses narrative aspects of oral language. Pupils' ability to recall the story is measured based on information content, sentence length, grammatical usage and independence. The assessment of narrative skills is a growing area of research. However, the Bus Story remains the most commonly used measure.¹⁹ The assessment will be conducted on a one-to-one basis by RAs trained in language assessment by an experienced child psychologist. These RAs will be blind to trial arm assignment.

Note that due to concerns over whether pupils targeted for participation in this trial will be able to engage with an RA when the assessment is administered (due to social anxiety), a pilot collection will occur in the summer 2018 school term. This pilot will seek to establish the best way to increase the rate at which children targeted for the intervention engage with the assessment. We will also estimate the correlation between scores obtained and EYFSP scores, if the latter can be obtained. In advance of the trial we will liaise with early year practitioners to establish an administration process that gives RAs the highest possible chance of engaging a child to complete the assessment. If the pilot raises serious concerns about children's ability to take part in the assessment, even after reasonable adjustments to administration procedure have been made, oral communication will become a secondary outcome measure.

Social Skills

An additional end-line secondary outcome will be social skill, as measured by the Social Skills scale of the Social Skills Improvement System (SSiS)²⁰. The SSiS contains three scales: the aforementioned Social Skills scale, a Problem Behaviours scale and an Academic Competence scale. As the intervention logic model most supports detecting a change in social skills, we will not administer the other two scales. The SSiS Social Skills scale assesses pupils' skills across the following subscales: communication, cooperation, assertion, responsibility, empathy, engagement and self-control.

SSiS is the most commonly used social skills assessment for young children, is standardised, and has been used in prior EEF evaluations.²¹ We chose to use SSiS, over an equally popular instrument, the Strengths and Difficulties Questionnaire (SDQ) because it is more thorough and in-depth than the SDQ. We also felt that the SSiS was a better tool to assess the types of behaviours we would expect to change as a result of participating in the programme (e.g., communication, engagement), whereas the SDQ was more oriented to identify problematic behaviours (e.g., conduct problems, peer relationship problems). The questionnaires will be delivered to teachers electronically. As with all measures of social skills at this age, this must

¹⁸ <http://www.talkingpoint.org.uk/slts/assessment-children-slcn/expressive-language-assessments>

¹⁹ Dockrell, J. E. (2001). Assessing language skills in preschool children. *Child Psychology and Psychiatry Review*, 6(2), 74-85.

²⁰ <https://www.pearsonclinical.com/education/products/100000322/social-skills-improvement-system-ssis-rating-scales.html>

²¹ Centre for Effective Education, Queen's University Belfast. (2016). Evaluation Protocol: Zippy's Friends. London, United Kingdom: Education Endowment Foundation. Retrieved from: https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Protocols/EEF_Project_Protocol_Character_Zippys_Friends_protocol.pdf.

be completed by the child's teacher and thus cannot be blind to trial arm assignment.

Creative 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. For this reason, we consider pupils' self-perception of ability to generate and use ideas in their school work as a secondary outcome measure. To measure this, we will use an adapted version of the ideation sub-measure of the writing self-efficacy measure proposed by Bruning et al. (2013), with significant simplification of language to make it appropriate for this age group (the original measure was designed for secondary school pupils).²² This approach has been taken to provide some scope for comparison with other trials being conducted at the same time (evaluation of Young Journalist Academy, Power of Pictures and Craft of Writing) in which we will also examine this subscale as part of the wider measure of writing self-efficacy. This measure will be captured using three, three-category Likert scale items asked by RAs after completion of the Renfrew Bus Story measure.

Analysis plan

Primary analysis

We will estimate the effect of the trial on reading attainment using a linear model on pupil-level data with treatment arm indicators, strata indicators, and a baseline covariate. As different versions of the PIRA test will be used for the Year 1 and 2 cohorts, raw scores will be standardised to have a mean of zero and a standard deviation of one, prior to combining cohorts for the purpose of analysis.

Our baseline covariate will be the child's EYFSP composite score for four learning goals:

1. understanding;
2. speaking;
3. reading; and
4. writing.

These goals were selected as they are most closely linked to reading. Past research found that neither the total EYFSP score nor the score for personal, social and emotional development are correlated well with future attainment.²³ This aggregate score will range from 4 to 12. To check for robustness, analysis will be performed using raw PIRA test scores for each year level separately.

To assess the effect of the intervention on oral communication we will estimate a linear model of Renfrew Bus Story test information scores on treatment assignment, strata indicators, and a baseline covariate consisting of EYFSP aggregate scores across the following learning goals:

1. listening and attention;
2. understanding; and
3. speaking

As above, the goals were selected as they are most closely linked to oral communication skills. This score will range from 3 to 9.

²² Bruning, R., Dempsey, M., Kauffman, D., McKim, C. & Zumbunn, S. (2013) Examining Dimensions of Self-Efficacy for Writing. *Journal of Educational Psychology*, 105(1), 25-38

²³ Snowling, M. J., Hulme, C., Bailey, A. M., Stothard, S. E., & Lindsay, G. (2011). Better communication research project: language and literacy attainment of pupils during early years and through KS2: does teacher assessment at five provide a valid measure of children's current and future educational attainments?. London: Department for Education.

The estimated impacts will be “intention to treat” (ITT) effects and will be reported with 95% confidence intervals. Effect sizes will be calculated using the Hedges’ g formula. 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.

Secondary analysis

Secondary analysis of social skills will use the SSiS scores produced by teachers, with a baseline covariate consisting of EYFSP aggregate score across the following learning goals:

1. self-confidence and awareness;
2. managing feelings and behaviour; and
3. making relationships.

This score will range from 3 to 9.

Secondary analysis of pupils’ self-perception of ability to generate and use ideas in their school work will again conform to the primary outcome estimate model, though no baseline covariate will be included.

Subgroup analysis

We will carry out a subgroup analysis to measure the impact of the intervention on everFSM pupils. Following EEF guidance, we will first test for an interaction of the treatment and everFSM status. If a significant interaction is found, we will estimate a separate model on the restricted sample of only everFSM pupils. This procedure will be conducted using a separate model estimated on the restricted sample of only everFSM pupils. This will be carried out for both our primary and our secondary outcomes.

Other

We will report the distribution of missing observations by treatment arm and explore whether baseline characteristics are balanced across trial arms.

We will estimate treatment effects across all three outcome measures for compilers using a Complier Average Causal Effect (CACE) analysis, using a pupil-level measure of compliance with the intervention. Compliance in this trial will be defined as having attended at least 16 of the 24 Speech Bubbles sessions. Attendance will be recorded by the drama practitioner and held centrally by the project team. It is important to note that we do not know the true minimal amount of compliance needed to generate a treatment effect, so the cut-off chosen for minimal compliance is our best estimate, which was defined in coordination with the delivery organisation. This analysis is likely to generate treatment effects that exceed those generated by ITT (unless the treatment is detrimental).

Definition of fidelity/on-treatment minimum

We outline below the fidelity measure and on-treatment minimum for Speech Bubbles 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.

The fidelity of this intervention will be measured at the pupil level, specifically the attendance of the pupil at the Speech Bubbles sessions. Of the 24 possible sessions, a pupil will need to attend a minimum of 16 to be considered on treatment. This equates to approximately 66% attendance. Attendance will be recorded by the drama practitioner and held centrally and will be collected by RAs during the assessments.

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 Speech Bubbles achieves positive outcomes for young people. In the first section, we outline the overarching implementation questions that will be explored across all projects, including Speech Bubbles. These cross-project similarities in delivery and in what they are aiming to achieve are outlined in the appendix.²⁴ We highlight, for each question, the dimension or factor affecting implementation it relates to, as specified in the guidance set out by the EEF.²⁵

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

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:
 1. Senior Leadership Team buy-in;
 2. Delivery of training and resources– 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;
 3. Delivery of the intervention – a) consistent across sites; b) 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 TAs? (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)

Speech Bubbles specific questions

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

1. What are the mechanisms that are taking place in the intervention and to what extent are they bringing about change? (Mechanisms)
2. Delivered by 7 separate partners across the country – to what extent is consistency ensured or the programme adapted? (Fidelity)
3. To what extent is the programme adapted by drama practitioners and schools? (Adaption)
4. What other support do the pupils access to support their communication in both control and treatment? Are pupils in control and treatment similar? (Programme differentiation)
5. To what extent does the intervention affect the targeted children’s classroom engagement and learning, particularly around engagement and communication? (Reach)
6. To what extent does the intervention affect the TA, their role in school and with the

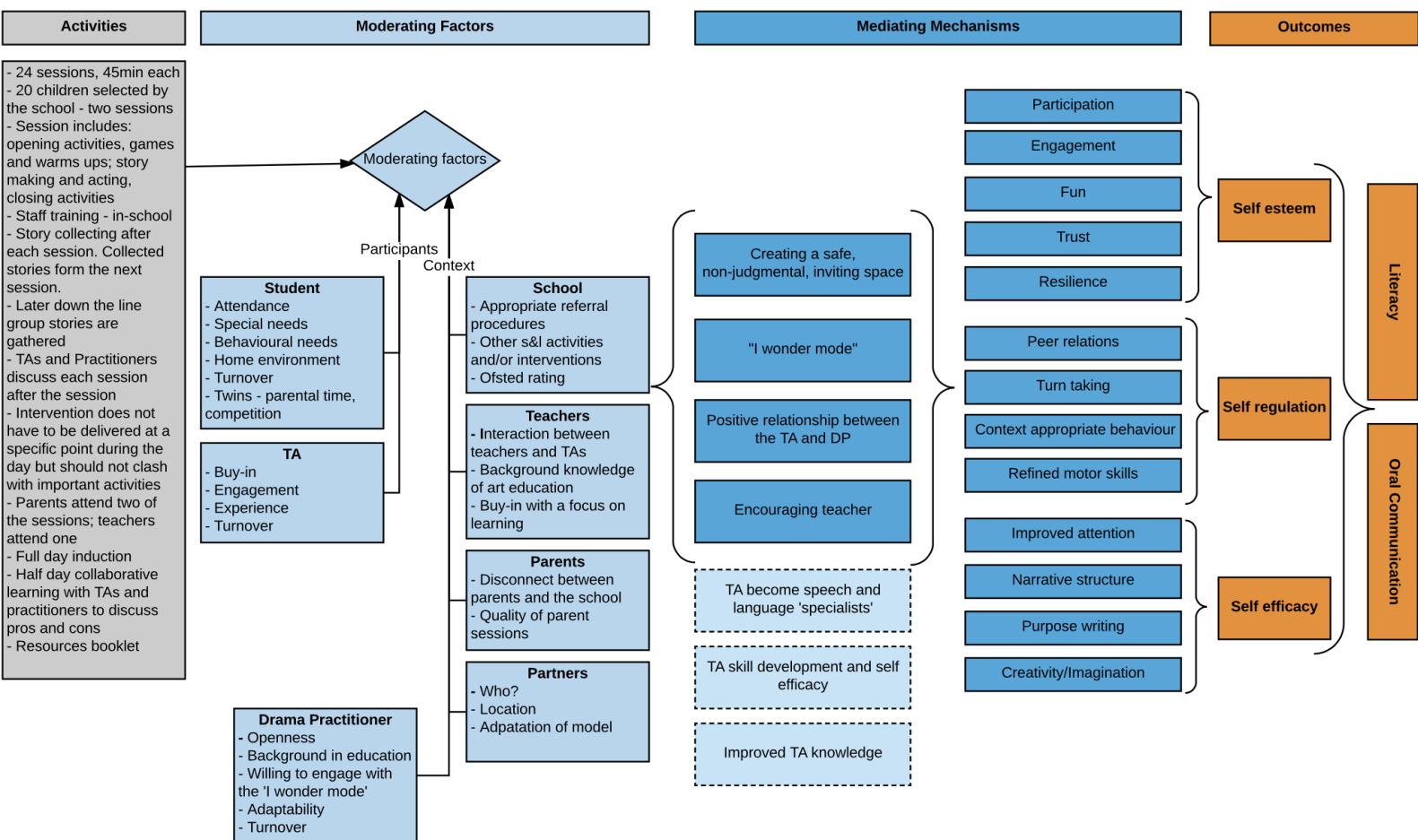
²⁴ For an overarching flow diagram of the programme similarities, please see appendix 2.

²⁵ 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.)*.

- class teachers? (Implementer characteristics and context)
7. To what extent do school facilities affect the intervention? (Implementation environment)

Logic Model

An IDEA workshop was held, utilising the TIDieR framework, to develop a logic model in collaboration with Speech Bubbles. 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 Speech Bubbles team to ensure we conduct the data collection when appropriate. In addition to the main project team input, Professor Andrew Burn, specialist in English, Media and Drama, Professor Gemma Moss, literacy specialist and Emerita Professor Sue Hallam, specialist in music education, will be invited to give feedback on the methods.

Observation of training

The IPE team will attend and observe the induction day training and one midpoint 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:

- TA interview after the observation
- Bubbles practitioner interview after the observation
- Observation of a Speech Bubbles session
- Informal interviews with children
- Interview with SLT
- Interview with classroom teacher

The schools will be sampled based on a range of characteristics such as geography, Ofsted rating and engagement (see defining fidelity above). 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 all 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, additional 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 is outlined in the table below.

Cultural Learning IPE Questions	Methods
<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"> 1. Senior Leadership Team buy-in; 2. Delivery of training and resources – 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; 3. Delivery of the intervention – a) consistent across sites; b) whether it appears to facilitate children’s engagement 	<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 TAs? (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>
<p>Speech Bubbles Questions</p>	
<p>What are the mechanisms that are taking place in the intervention and to what extent are they bringing about change? (Mechanisms)</p>	<p>Survey; Observation, Case studies</p>
<p>Delivered by 7 separate partners across the country – to what extent is consistency ensured or the programme adapted? (Fidelity)</p>	<p>Survey; Observation</p>
<p>What other support do the pupils access to support their communication in both control and treatment? Are pupils in control and treatment similar? (Programme differentiation)</p>	<p>Administrative data; Survey; Case studies</p>
<p>To what extent is the programme adapted by drama practitioners and schools? (Adaption)</p>	<p>Case studies; Survey</p>
<p>To what extent does the intervention affect the targeted children’s classroom engagement and learning, particularly around engagement and communication? (Reach)</p>	<p>Survey; Case studies</p>
<p>To what extent does the intervention affect the TA, their role in school and with the intervention teacher? (Implementer characteristics and context)</p>	<p>Survey; Administrative data</p>
<p>To what extent do school facilities affect the intervention? (Implementation environment)</p>	<p>Administrative Data; Survey</p>

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 software 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 Speech Bubbles team are yet to specify exact timings for their programme delivery across the three locations. We can therefore consider this an indicative schedule of events across the academic year of 2018-19.

Date	Item
Autumn Term 2018	Observation of first training session
	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
	Conduct in-school case studies
	Collection of fidelity data to inform case study sampling
	Finalise sampling strategy
Summer Term 2019	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 financial costs of implementing the intervention. This includes anything which the school needed to pay for beyond business as usual.

The cost estimates will make use of information from the project team (particularly with regard to 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 supply cover for teachers to attend training, but also to prepare for delivery, will be reported separately from the financial costs, for example printing costs of materials. Any costs in terms of prerequisites will also be considered. 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. Any costs associated purely with the evaluation will be excluded.

Ethics and registration

Ethical approval has been sought following UCL Institute of Education staff ethics approval procedure. It was approved on 11 December 2017.

Personal data for this trial will be processed under the legitimate interests provision of the GDPR (Article 6(1)(f)). Nevertheless, parents will be provided with information explaining their right to withdraw their child's data from the trial if they object to this processing of their child's data. This use of data has been allocated the following UCL Data Protection Registration Number: Z6364106/2017/11/56.

This trial protocol has been pre-registered at www.controlled-trials.com, and assigned an International Standard Randomised Controlled Trial Number (ISRCTN) of 14448319.

Personnel

Delivery team:

- Adam Annand (London Bubble)
- Amelia Bird (London Bubble)

Evaluation team:

BIT

- Pantelis Solomon (Principal Investigator)
- Kim Bohling

- Matt Barnard
- Florentyna Farghly
- Fabian Gunzinger
- Johanna Frerichs
- Alex Manby
- Millie Devereux

UCL Institute of Education

- Jake Anders (Principal Investigator)
- Dominic Wyse
- Gemma Moss
- Andrew Burn
- Nikki Shure
- John Jerrim
- Susan Hallam

Responsibilities

Outcome measures administration and collection - BIT

Design of the trial

- sample size calculations - BIT
- refinement of randomisation approach - BIT

Delivery of the intervention

- recruitment of schools - London Bubble and partners
- session delivery - London Bubble and partners

Data collection

- Collection of pupil data - BIT
- Outcome measure collection (RA recruitment and coordination) - BIT
- Linking of UPN to NPD - IoE
- Data for process evaluation - BIT

Impact analysis - BIT (lead) and UCL

Qualitative analysis - BIT (lead) and UCL

Risks

The data security policies of UCL and BIT and the Data Sharing Agreement between BIT, UCL, and London Bubbles will be added to this protocol once drafted and approved.

Issue/risk	Risk level	Action to address issue/reduce risk
Dropout/non-compliance of settings	Medium	<p>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 MoU). Keeping schools informed of progress and providing reminders of next steps will be important for engagement. The project team will also be asked to monitor changes in key personnel to ensure ongoing commitment.</p> <p>Minimising the data collection burden on schools will also be important for retention. We will also randomise only after schools have given parents the opportunity to opt-out and have provided the necessary pupil data.</p> <p>Recruiting the target of 25 schools will also provide a buffer with respect to sample size reduction.</p>
Difficulty in collecting pupil data prior to randomisation	Medium	<p>We will work closely with delivery teams and maintain regular contact. A school recruitment timetable, which includes a communication schedule, will be shared and agreed with the delivery partners. As part of this, delivery partners will be asked to send a weekly email, or to update a shared spreadsheet with contact details of recruited schools.</p> <p>Pupil data will be submitted directly to BIT, who will screen each data set as it comes in to check for any incomplete or incorrectly entered data, to ensure a school is eligible for randomisation.</p> <p>The school recruitment timetable builds in sufficient time to follow up with schools who have either not returned their data on time or have returned incomplete data to ensure that randomisation is not delayed.</p>
Difficulty recruiting schools	Low to medium	<p>We are confident that the project team will convey the importance of the evaluation to settings and the value to them of taking part. As the trial is individually randomised children at each recruited school will have the opportunity to take part in the intervention.</p>

Contamination	Medium	<p>Communications from both the project and evaluation team will stress that children assigned to the control condition cannot be given access to Speech Bubbles sessions, even in the event those assigned to treatment cease to attend sessions.</p> <p>The school MoU will also be explicit on this point.</p> <p>Practitioners delivering Speech Bubbles sessions in schools will also be asked to check those attending are listed in the treatment group list. In the event a school wishes to replace a child who has ceased to attend Speech Bubbles sessions, that child will need to be drawn from beyond the control group, with the project team notified (who will then in turn notify BIT).</p>
Objection to NPD linking	Medium	<p>We believe this processing of personal data is justified under the legitimate interests basis of GDPR (Article 6(1)(f)). Nevertheless, we will inform parents of the right to withdraw their child from the processing of their data. We do not anticipate high or non-randomly varying levels of withdrawals.</p>
Missing Outcome Data	Medium	<p>For directly collected assessments, attrition is a potential risk. BIT will ensure schools and RAs understand the need to collect end-line measures for as many students as possible.</p> <p>Schools will also be contacted sufficiently far ahead of the end-line primary data collection window to ensure we arrive at a convenient time for RAs to visit and run the PIRA tests (in order to avoid weeks or days in which large numbers of pupils are likely to be absent). Upfront notice will also give school teachers ample time to complete SSiS surveys.</p> <p>RAs will report the number of children not able to sit the PIRA to the BIT evaluation team after each visit. If the rate is high (>5% of sample) a project coordinator will contact the school to determine whether an additional testing day is appropriate and/or feasible. BIT will provide regular updates to the EEF about testing completion and attrition.</p> <p>To ensure PIRA response papers are not lost in the postal system, they will be couriered to the test publisher for marking. Once marking is complete, the test publisher will then send BIT an electronic record of marks (over a secure service) and courier the hard copy papers themselves.</p>

		We will use a version of the PIRA test set at a lower difficulty standard, in order to enable more children to engage with it.
Parent and teacher concern about 'over-testing'	Low	Communications to schools (during recruitment) and parents (when providing the opportunity to opt-out) should emphasise that only two assessment will be taken by children due to this study (PIRA and Renfrew Bus Story).
Problematic randomisation	Low	The randomisation will be conducted at the individual level and as such should be relatively simple to implement. As randomisation is conducted by school, BIT will keep track of the absolute numbers of year 1 and year 2 children in each trial arm, and the number of children who have been opted-out, to ensure balance is achieved.
Treatment variation	N/A	We view this not so much as a risk but as the reality of implementing such an intervention. The impact estimates (ITT) 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. As BIT and UCL are joint evaluators, there is a relatively large pool of staff with experience in education evaluation who could substitute for members of the team, should this be necessary.

Timeline

Date	Activity
October 17 - March 18	School recruitment (London Bubbles)
January - March 18	MoU signing and ensuring children are referred by teachers (London Bubbles)
April - May 18	Distribution of opt-out forms to parents of referred children (BIT)
Late May 18	Final date of return of opt-out forms before schools send pupil

	data to evaluators (BIT)
June - mid July 18	Rolling randomisation as data upload and opt-out process completion confirmed (BIT)
September 18 - July 19	Intervention delivery (London Bubbles)
September-October 18	Observe training (BIT)
October 18	NPD application (UCL) and IPE baseline survey (BIT)
February 19	Observe second training (BIT)
March-April	Conduct sampling for case studies (BIT)
May - July 19	End-line (PIRA and Renfrew Bus Story test) administered by RAs (BIT), SSiS completed by teachers (BIT), and case studies for IPE conducted (BIT)
July	End-line IPE survey (BIT)
July 19	Marking of PIRA end-line assessments (Hodder, contracted by BIT)
September - December 19	Analysis and report writing (BIT and UCL)

Annex 1: Speech Bubbles referral guidance

There are three referral categories:

Children who:

Lack confidence in communicating:

This includes children who are selectively mute and those with English as an additional language.

Have difficulty organising thoughts and then communicating them:

This includes children who may not respond appropriately to what is being said.

Have poor attention, poor listening:

This includes children with a low level of engagement with classwork and difficulty developing positive peer to peer relations.

However:

The programme is not designed to address issues that would require referral for individual speech and language therapy and may not be effective for children with complex emotional and behavioural concerns.

You may also want to consider if the children you refer are undertaking other interventions and weigh up the costs and benefits of time out of the classroom.

'The referral guidance is appropriate for the project and addresses all the key areas of need.'
S Bojic-Macintosh BSc (Hons) MRCSLT, MASLTIP Highly Specialist Speech and Language Therapist.

Annex 2: Cultural programmes IPE similarities

Similarities across projects

The logic models from the five 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 1).

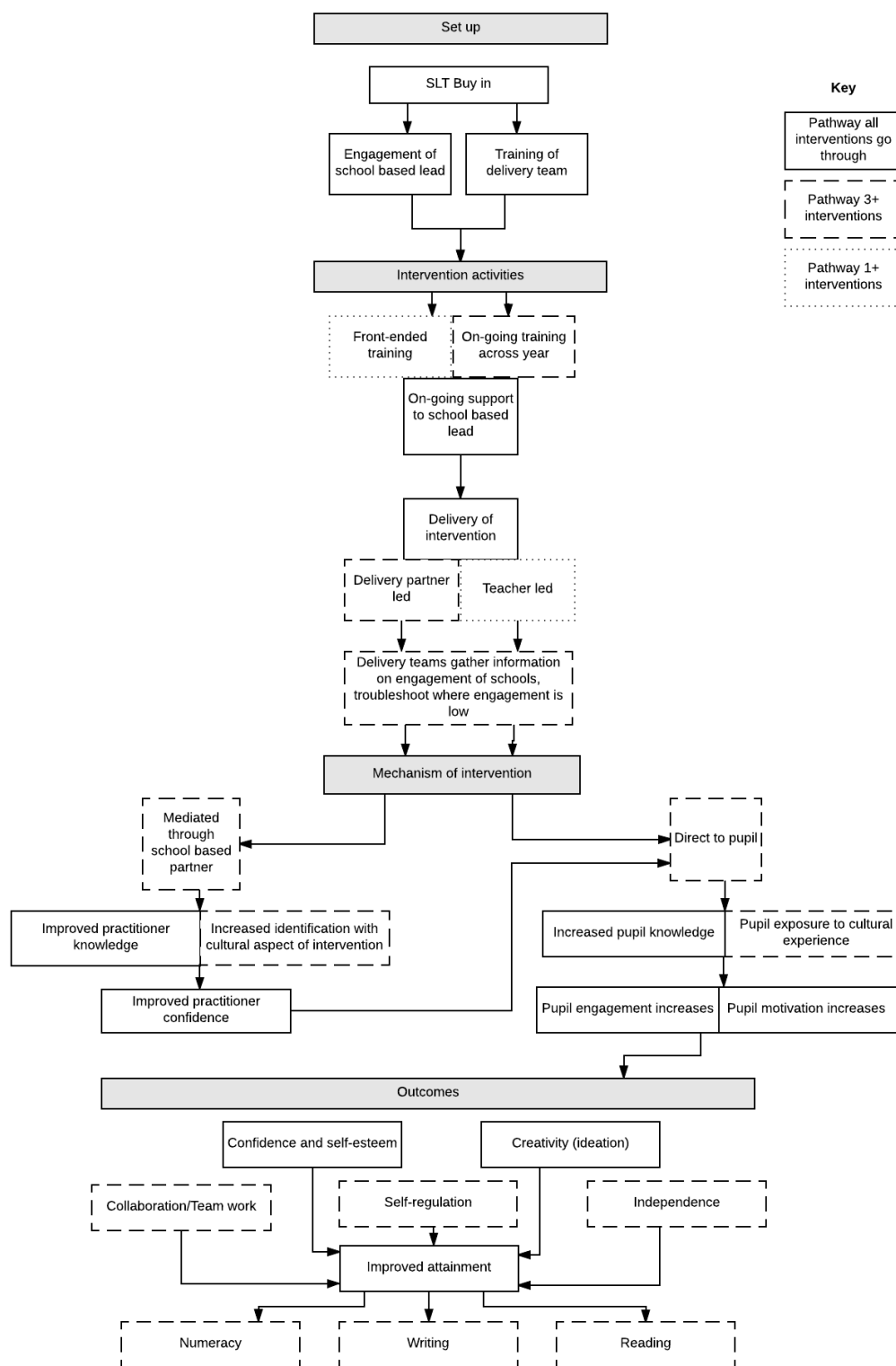


Figure 1 Amalgamated logic model of the five interventions

From Figure 1, we can see that the following are standard across all five interventions:

Implementation Similarities

1. Senior leadership buy-in
2. On-going (yet varied) support from delivery team staff - relationship with school, and teachers or TAs.
3. Training days for teachers or TAs.
4. Delivery teams gather information which helps them to 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 five 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 four 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 four 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