



## **SPEECH BUBBLES**

Evaluation Report

September 2021

Kimberly Bohling, Lev Tankelevitch, Andriana Vinnitchok,  
Matthew Barnard, Jake Anders, Nikki Shure, Dominic Wyse



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

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

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

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

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

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

- 📍 Jonathan Kay  
Education Endowment Foundation  
5th Floor, Millbank Tower  
21–24 Millbank  
SW1P 4QP
- ☎ 0207 802 1653
- ✉ [jonathan.kay@eefoundation.org.uk](mailto:jonathan.kay@eefoundation.org.uk)
- 🌐 [www.educationendowmentfoundation.org.uk](http://www.educationendowmentfoundation.org.uk)

# Contents

About the evaluator .....	4
Executive summary .....	5
Introduction .....	7
Methods.....	18
Impact evaluation results .....	40
Implementation and process evaluation (IPE) results .....	57
Cost .....	70
Conclusion .....	72
References .....	77
Appendix A: EEF cost rating.....	79
Appendix B: Security classification of trial findings.....	80
Appendix C: Effect size estimation .....	83
Further Appendices .....	84

## About the evaluator

The project was independently evaluated by a team from University College London (Dr Jake Anders, Dr Nikki Shure, Professor Dominic Wyse, Professor John Jerrim, Professor Gemma Moss, Professor Andrew Burn) and the Behavioural Insights Team (Dr Matthew Barnard, Kimberly Bohling, Dr Lev Tankelevitch, Andriana Vinnitchok).

The lead evaluators were Kimberly Bohling and Dr Matthew Barnard.

BIT contact details:

Kimberly Bohling

Behavioural Insights Team

4 Matthew Parker Street

London SW1H 9NP

Email: [info@bi.team](mailto:info@bi.team)

## Executive summary

### The project

The Speech Bubbles (SB) intervention is a Key Stage 1 (KS1) drama and storytelling intervention aimed at supporting children's communication skills, confidence and wellbeing. The intervention was targeted at pupils aged 5–7 in KS1 (Year 1 or Year 2) in primary schools in England, who had been identified by their teachers as having difficulties with, or lacking confidence in, communicating, or having poor attention and listening skills.

The SB intervention consisted of weekly 45-minute sessions, which were delivered to two mixed-year groups of 10 pupils during normal teaching hours for 24 weeks over three school terms. Each session followed a clear and repeated routine centred around the telling and re-enacting of stories that are told by the children. The sessions included activities that support expressive language, receptive language and turn taking. Each SB session is delivered by a drama practitioner (DP), who is recruited and trained in the SB approach by London Bubble, and a teaching assistant (TA) who works at the respective school in which the SB sessions take place.

The trial was a two-arm individually randomised controlled trial (RCT). 1006 pupils across 26 schools were recruited to participate, of whom 40 pupils from each school were randomly allocated to either the intervention-arm (who received the programme) or the control arm, at a 50 : 50 ratio within each school. Between June 2018 and July 2019, the trial tested the efficacy of the SB programme on reading attainment, oral communication, creative self-efficacy and social skills.

An implementation and process evaluation (IPE) was conducted to complement the findings from the impact evaluation. The IPE focused on implementation, delivery and perceived impact of the intervention. IPE data collection consisted of collecting case study data from six schools, alongside complementary surveys for all schools taking part in the trial. Each case study involved the following methods: a semi-structured interview with the SB facilitators (i.e., the TA and DP), a Year 1 or Year 2 teacher, and a member of the senior leadership team (SLT); an observation of an SB session; and informal discussions with participating pupils. This evaluation was jointly funded by the EEF and the Royal Society of Arts.

#### Key conclusions

1. Pupils who received the Speech Bubbles intervention had, on average, lower reading attainment scores (equivalent to one month's less progress) as compared to children in the control group. Pupils who received Speech Bubbles did not show any substantial difference in oral communication skills as compared to pupils in the control group. These are our best estimates of impact which both have a high security rating. However, as with any study, there is uncertainty around the result: the possible impact of this programme on both reading attainment and oral communication ranges from three months less progress to positive effects of two additional months of progress.
2. The impact evaluation did not find evidence the Speech Bubbles intervention had any effect on creative self-efficacy or social skills. The intervention did not have a differential impact on reading attainment or oral communication skills by eligibility for free school meals.
3. Members of the SLT, TAs and drama practitioners in the case study schools generally reported feeling positive about the intervention and reported noticing positive changes in pupils' oral communication skills, self-efficacy, and self-regulation. However, the perceived impacts on pupils' literacy were limited.
4. Overall, fidelity to the delivery model and dosage were high, based on data gathered through interviews, surveys, observations, and delivery records.
5. The key factors identified that influenced implementation and delivery of the intervention were: (i) the need to have SLT support to secure required resources (e.g., the same room each week); (ii) SB facilitators acting as role models and partners in delivery; and (iii) the characteristics of the pupils and associated group dynamics during the SB sessions.

### EEF security rating

These findings have a high security rating. The trial was a well-designed two-arm randomised controlled efficacy trial which tested whether the intervention worked under developer-led conditions in a number of schools. Pupils in SB classes were similar to those in the comparison classes in terms of FSM and EAL eligibility. The trial was not as well-

powered as originally intended due to of the inability to access baseline measures as planned (discussed in the Introduction), and a lower number of pupils at the analysis stage: 18.4% and 19.4% [respectively, for Progress in Reading Assessment (PIRA) and Renfrew Bus Story (RBS), the two primary outcomes] of those who started the trial were not included in the final analysis due to pupils changing schools, pupil absence and schools not providing test data.

## Additional findings

Pupils receiving the SB programme made, on average one fewer month's progress than those in the control group equivalent in reading attainment scores. The intervention was found to have no positive effect on oral communication skills for pupils in KS1. This is our best estimate of impact, which has a high security rating. However, as with any study, there is some uncertainty, with the possible impacts on both reading and oral language outcomes ranging between negative effects of three fewer months' progress and positive effects of two additional months' progress.

Compliance with delivery frequency and fidelity to the programme was very high. Of the pupils for whom attendance data is available (88% of the treated sample), 87% met the minimum compliance threshold of attending 16 of 24 SB sessions, and pupils attended a median of 22 sessions. Data gathered through interviews, surveys and observations provided evidence that fidelity to the general structure of the sessions was maintained with minimal modifications.

A high level of compliance is an indication that teachers, TAs and schools felt positively about the SB programme. Data from the school survey found that 86% of staff co-facilitators felt that the intervention had a positive impact on their understanding of how children learn. A large majority of TAs who completed the survey perceived that SB improved pupils' confidence, social skills and communication. However, the impact evaluation did not find evidence of an effect on reading attainment and oral communication skills, and a much lower proportion of TAs in the survey reported that they felt the intervention had a positive impact on literacy (38%).

It is crucial to note that the intervention was delivered in a real-world context and all trial schools surveyed indicated they had other support available for pupils with speech and language challenges. Ninety percent of surveyed schools indicated that in-class and out-of-class adult support was routinely available (90% and 71% of schools, respectively). Within this context, the addition of SB may not have added much beyond the business-as-usual support offer. Additionally, the eligibility criteria for the targeted intervention were fairly broad, which led to the selection of a mixed group of pupils who had speech and language difficulties for a wide variety of reasons. Some SB co-facilitators reported that the combination of pupils with different dispositions made it sometimes more difficult to manage the group dynamics of the SB session. It may therefore be worth revising the intervention and undertaking further research on future iterations of the intervention that are targeted more specifically to pupils with particular difficulties.

## Cost

The average cost of SB per school is around £110 per pupil per year when averaged over three years.

## Impact

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

Outcome / group	Effect size (95% confidence interval)	Estimated months' progress	EEF security rating	No of pupils	p-value	EEF cost rating
Reading attainment (Progress in Reading Assessment: PIRA)	-0.05 (-0.2, 0.1)	-1		411 (intervention group), 410 (control group)	0.93	£ £ £ £ £
Oral communication (Renfrew Bus Story test: RBS)	-0.04 (-0.19, 0.11)	0		405 (intervention group), 406 (control group)	> 0.99	£ £ £ £ £

## Introduction

### Background

This evaluation is part of a round of funding between the Education Endowment Foundation (EEF) and the Royal Society of Arts (RSA) to test the impact of different cultural learning strategies in English schools, entitled 'Learning about Culture'.<sup>1</sup> The aim is to improve the evidence base around arts-based education programmes. It consists of five programmes: two in Key Stage 1 (KS1) (Years 1 and 2) and three in Key Stage 2 (KS2) (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> The programmes have been supported by Arts Council England.

The background for the study is that a focus on increasing attainment in literacy and numeracy has been criticised for leading to a marginalisation of art, music and cultural studies in English schools (Warwick Commission, 2015). The UK Government's Culture and Sport Evidence Review (Newman et al., 2010), which summarised much of the observational and qualitative research in this area, showed pupil participation in cultural learning programmes (from piano training to theatre-based drama projects) to be correlated with higher levels of achievement in mathematics and literacy / English in both primary and secondary school. The review also linked participation in cultural learning programmes to faster language development in the early years, and improved cognitive ability. Additionally, large cohort observational studies in the US have suggested that the mathematics and literacy gains to cultural participation are particularly large for pupils from low-income groups (Catterall, 2009; Catterall et al., 2012). This evidence suggests that cultural learning correlates with academic attainment; however, the causal nature of this relationship remains unclear – a key motivating factor for the 'Learning about Culture' programme.

The Speech Bubbles (SB) intervention is targeted at pupils assessed by teachers as having below-expected communication and social skills, and it aims to improve these skills by providing pupils with weekly, group-based creative drama sessions. There is some evidence to suggest that drama and other arts-based programmes similar to SB can improve academic attainment across several language-related areas. Notably, the majority of evidence consists of non-experimental studies, thereby preventing causal inference, or experimental studies with less robust designs, such as non-randomised treatment assignment or small samples (See & Kokotsaki, 2015; Lee et al., 2020). Nevertheless, large-scale reviews of evidence of the effect of arts education on academic achievement identified a potential causal link between classroom drama and an improvement in a variety of verbal areas (Hetland & Winner, 2001; See & Kokotsaki, 2015; Lee et al., 2020). Effects were found in domains of written understanding and recall of stories, and 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 practised during drama sessions (Podlozny, 2000; Hetland & Winner, 2001; Lee et al., 2020). Some evidence also points to a link between drama-based programmes and improved oral and written communication and social skills, specifically among pupils with language and communication difficulties (e.g., those with learning or developmental disabilities, autism spectrum disorder, or attention deficit hyperactivity disorder: see Adamek et al., 2017). However, there is a dearth of high-quality randomised controlled trials (RCTs) rigorously testing the impact of such interventions, particularly among pupils with below-expected communication and social skills (See & Kokotsaki, 2015; Adamek et al., 2017; Lee et al., 2020).

At the time of funding the evaluation (2016–17), SB was running in 44 schools across the country, and had been the subject of three research projects (Barnes, 2014; 2015; Price & Ansong, 2016). The programme was developed by the London Bubble Theatre Company, with professional support from speech therapists, educational psychologists and Southwark Pupil Development Centres.

In 2013, the Shine Trust funded a mixed-methods evaluation, including a pre/post comparison of the SB programme (Barnes, 2015). At the time, the programme was running in 29 schools in disadvantaged areas of London and North West England. Multiple sources of data were used in the evaluation, including school records, the reports of theatre

---

<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 the Supplemental Appendices.

practitioners, parents, class teachers and speech therapists; and a team of independent teachers were also consulted. Feedback from teachers suggested that over 80% of SB children showed improvement beyond expectation in their learning, speaking and listening skills, and in over 50% of these children the progress was either clear or striking. In addition, teachers reported that 85% of participating children showed progress in their emotional and conduct behaviour.

A quasi-experimental 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 SB programme made very good progress relative to an unmatched comparison group with similar support needs for speech and communication development (Price & Ansong, 2016). Children in the treatment group showed significantly faster improvement in their speech, 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.

This evaluation provided an opportunity to build on this evidence and explore the impact of SB using robust experimental methods that allow for a causal attribution of any measured change, together with an implementation and process evaluation (IPE), to understand how SB was delivered within schools, the barriers and facilitators to implementation, and fidelity.

The impact evaluation consisted of an individually randomised controlled trial (RCT). Year 1 and Year 2 pupils with poor communication skills (as identified by teachers using SB referral guidance, described below) were randomly allocated to either receive the SB programme (treatment arm), or continue with their curriculum as normal (control arm). Pupils in the two arms were compared on reading attainment and oral communication as two primary outcomes, and on social skills and creative self-efficacy as two secondary outcomes. Assuming successful randomisation, an RCT is the least biased method to estimate intervention effects and make causal inferences.

In the IPE, multiple sources of data were collected and triangulated from six case study schools to answer the research questions. This consisted of semi-structured interviews with four staff members at each case study school, unstructured interviews with participating children, an observation of an SB lesson, a survey for teaching assistants (TAs), review of intervention manuals and guidance, and collection of administrative data.

## A note on protocol deviations related to data sharing

We note upfront that it has been necessary for the analysis of this trial to deviate substantially from our initial plans set out in the project protocol and statistical analysis plan (SAP).<sup>3</sup> These stem from issues in accessing the baseline data that we expected to be able to obtain from the Department for Education (DfE)'s National Pupil Database (NPD). During the implementation of the trial, the DfE changed the way in which data from the NPD are made available to researchers, switching from providing extracts that can be used alongside project data within evaluator's own secure computing systems to requiring access within the Office for National Statistics (ONS) Secure Research Service (SRS). In turn, this means that it is now necessary for project data to be uploaded to the SRS. Given that this project data is considered personal data over which we as evaluators are data controllers, this requires the conclusion of an appropriate data sharing or processing agreement between the evaluator and the DfE and/or the ONS in order to provide legally required reassurance by the DfE/ONS about the treatment of personal data over which the evaluator is controller. This implication of altering their processes appears not to have been fully planned for by the DfE and ONS and, as such, attempting to conclude such an agreement that does not appear to have precedents has been subject to extended negotiations and delays between the evaluators and DfE/ONS, which we understand to have been severely exacerbated by additional workload for these organisations due to the COVID-19 pandemic. In the interests of completing these evaluations, and after discussion with the EEF and project teams, the decision was made to proceed with the analysis, with deviations from protocol flagged as we move through the methods section. The main change is the use of pupils' FSM and EAL status as baseline measures, instead of Early Years Foundation Stage Profile (EYFSP) scores. These deviations were

---

<sup>3</sup> Speech Bubbles Evaluation Protocol; Speech Bubbles Statistical Analysis Plan (SAP).

agreed with the EEF and the developers ahead of conducting the analysis. Beyond issues inherent in deviating from pre-registered protocol, the main implication for the analysis is a reduction in the statistical power relative to expectations. This means that the impact evaluation is not able to statistically detect effects as small as expected in the SAP.

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

## Intervention

The SB programme is designed to improve children's communication and social skills by providing them with weekly creative drama sessions. The description below follows the Template for Intervention Description and Replication (TIDiER) checklist.<sup>4</sup>

- 1 **Brief name.** Speech Bubbles (SB)
- 2 **Why (rationale/theory).** The Speech Bubbles (SB) intervention aims to improve children's communication and social skills by providing them with weekly group-based creative drama sessions (see Figure 1 for a logic model of the intervention). These sessions include activities that support receptive and expressive language (how a pupil understands and uses language, respectively), turn-taking and turn-giving, and an understanding of story structure. As the core focus of the intervention is storytelling and working with narrative structure, activities which are relevant to reading, it is also thought that it may improve pupils' reading skills. SB is based on the Helicopter Stories pedagogical approach, which uses storytelling and story acting to engage children and enable them to collaborate, create narratives and draw connections together with adult facilitators and classmates (Lee, 2015).
- 3 **What (materials).** Resources and materials<sup>5</sup> for the intervention include (see *Activities* in Figure 1):
  - *Speech Bubbles resource pack* describing the rationale, approach, structure and activities;
  - *Narrative stimulus pack* to support the narrative of the stories;
  - *Emotional faces stimuli* to support children who struggle to convey emotions via language;
  - Agendas outlining the focus of parent sessions and continuing professional development (CPD) sessions (see Supplemental Appendices for CPD induction day details).
- 4 **What (procedures).** Each SB session follows a clear and repeated routine centred around the telling and re-enacting of stories that are told by the children (see *Activities* in Figure 1). The sessions include activities that support expressive language, receptive language and turn-taking. Sessions are led by the drama practitioner (DP), who is an external facilitator, with assistance from a teaching assistant (TA) from the school.

Sessions typically begin with opening activities, that are repeated across sessions, and which aim to create a structured, collaborative and fun environment (e.g., in the 'name in the bucket' game, children project their name into an imaginary bucket). Next, sessions include warm-up activities which aim to stimulate the

---

<sup>4</sup> <https://www.bmj.com/content/348/bmj.g1687>

<sup>5</sup> London Bubble. *Speech Bubbles* [online] is available at: [https://www.londonbubble.org.uk/parent\\_project/speech-bubbles/](https://www.londonbubble.org.uk/parent_project/speech-bubbles/) (accessed: 30 November 2020), for all resources relating to the Speech Bubbles intervention.

children's senses and prepare them as individuals and as a group for the core activity (e.g., in the 'bubbles' game, children blow an invisible bubble, step inside, clean the walls so they can see outside, and float up).

The core activity of each session is story-making, in which children create, share, explore and develop stories and their characters. This consists of (1) preparatory exercises, in which children practise key characters, events or pieces of scenery in the day's story; (2) marking out of the 'story square', in which the DP uses tape to mark the space where the story will be acted out; (3) storytelling and acting, in which the DP reads out the day's story, and calls children to act characters and setting; and (4) 'whoosh!', in which story participants are instructed mid-performance to leave the story square and allow a new group to enter.

Closing activities aim to provide a positive end to the session and prepare children to return to class (e.g., 'washing off characters', in which children describe characters that they have played during the session, thereby emphasising the 'pretend' aspect of the SB activities). After the sessions close, one or two children stay behind to tell the DP or TA their story for next week, which is written down verbatim (the 'story collection' process). In the following session, this story is read out to the group and re-enacted as described above. Lastly, after children leave each session, the DP and TA assess all of the children for development in receptive, attending and expressive skills. Using a standard form, each individual child is rated on a scale of 1–5, where 1 is 'unable to' and 5 is 'excellent' (or 'A' if the child is absent) for receptive skills ('Understanding and making sense of what people say and do.');

attending skills ('Giving and taking turns. Listening.');

and expressive skills ('Being able to communicate physically and verbally'). For each child, there is also a dedicated space to add further comments or indicate if any actions are needed. Informal evaluation at the group level is also done by making notes in a separate document on how the session went, overall.

To enable pupils to become comfortable with the SB programme, activities are introduced gradually. In the first session, it is common that the 'story square' is introduced by using a pre-prepared children's story. At the end of the first session, it is also common for the DP and TA to act out the 'story collection' process, so as to model it for the children. Other than this, there is no prescribed sequence for the sessions.

- 5 Who (recipients).** SB is targeted at KS1 (Year 1 and Year 2) pupils who display at least one of the following:
- a** *Lack of confidence in communicating* – this includes children who are selectively mute and those with English as an additional language (EAL);
  - b** *Difficulty organising thoughts and communicating them* – this includes children who may not respond appropriately to what is being said;
  - c** *Poor attention and poor listening skills* – this includes children with a low engagement with classwork and difficulty developing positive peer-to-peer relations.

Such pupils are identified and referred by their teachers, on the basis of guidance from the SB programme (see *Moderating factors* in Figure 1).<sup>6</sup> Referral guidance indicates that 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.

- 6 Who (implementers).** SB is delivered by drama practitioners (DPs) and teaching assistants (TAs). Teachers support the delivery of the intervention by providing ongoing feedback about pupils' progress and by facilitating pupil attendance (see *Moderating factors* in Figure 1). Parents are encouraged to attend two of the sessions, but this is not mandatory.

DPs are recruited by London Bubble and trained in the SB approach. Many have a drama school or university-level qualification in performance or applied theatre, and they are expected to come to the programme with a background in creative theatre-making with children. In addition, they are required to shadow other current SB practitioners (a minimum of two different practitioners, but duration varies) before delivering the programme themselves. DPs are also

---

<sup>6</sup> See Appendix F for Speech Bubbles referral guidance.

required to complete the Communication Trust online short course.<sup>7</sup> All DPs have experience working with children and young people, had an enhanced Disclosure and Barring Service (DBS) check within three years of the evaluation, and are given a safeguarding induction.

DPs lead on the planning of work, including plans for each term and session, and on the drama activity during each session. During each session, DPs promote the children's positive emotional engagement by using positive, open facial expressions; encouraging children to support each other's contributions; modelling a playful approach to activities; and other techniques aligned with the programme values. They support TAs in safeguarding the children, and together evaluate the sessions.

TAs are selected by each school; as children in a single SB cohort come from different classes, this may be the classroom TA for some children and a non-classroom TA for other children. TAs lead on knowledge of individual children and their safeguarding. They also lead on communication with schools, teachers and parents. TAs support the planning, delivery and evaluation of sessions with DPs. The same TA is expected to participate in every session, as consistency is an important aspect of the programme.

DPs, TAs and class teachers together attend a centrally held, six-hour induction day at the beginning of the intervention, led by London Bubble. One aim of the induction day is to introduce DPs, TAs and teachers to each other. A second aim is to introduce TAs and teachers to the SB programme, including its theoretical framework, activities and resources. This is done through modelling an SB session and discussing its approach and activities. School staff from schools who have previously done the SB programme are also present for the induction, and are able to provide context and advice for less experienced schools. A third aim is to enable DPs and TAs to discuss the requirements and practicalities of the sessions and to ensure mutual understanding between both parties. Finally, the day includes a discussion of the space requirements for the programme.

DPs, TAs and teachers also attend a centrally held, three-hour continuing professional development (CPD) session mid-way through the intervention. The session is led by London Bubble and includes extensive open discussions between DPs, TAs and other school staff. This is a collaborative learning opportunity to reflect on the progress of the intervention, any challenges and opportunities to learn from different schools. The focus of the session is to ensure that DPs and TAs understand the aims of the SB approach and the main features of the intervention. Secondly, the session aims to clarify whether schools are supportive of the intervention and the TAs delivering it, and whether and how they make time and space for the intervention within the curriculum. Lastly, it aims to clarify whether TAs feel confident in delivering the intervention material, their understanding of the DPs' expertise, and the learning and interaction between schools and DPs.

Additionally, all participating schools are offered an introductory taster session for all KS1 staff in their school, to help them understand the SB programme and approach. This is delivered as a 45-minute twilight session prior to the programme beginning implementation.

**7 How (mode of delivery).** Sessions are delivered face-to-face by DPs and TAs to two mixed-year groups of 10 pupils from Year 1 and Year 2.

**8 Where (setting).** Sessions are delivered in settings that are selected based on seven factors:

- **Regularity:** the same space each week so it feels familiar and safe;
- **Size:** large enough for 10 children to move around, but not so big that they feel lost (e.g., a classroom);
- **Volume:** as quiet as possible to allow the children to focus on the session (e.g., minimises echoes);
- **Privacy:** somewhere that other people do not walk through or look into;
- **Light:** natural light is preferred;

---

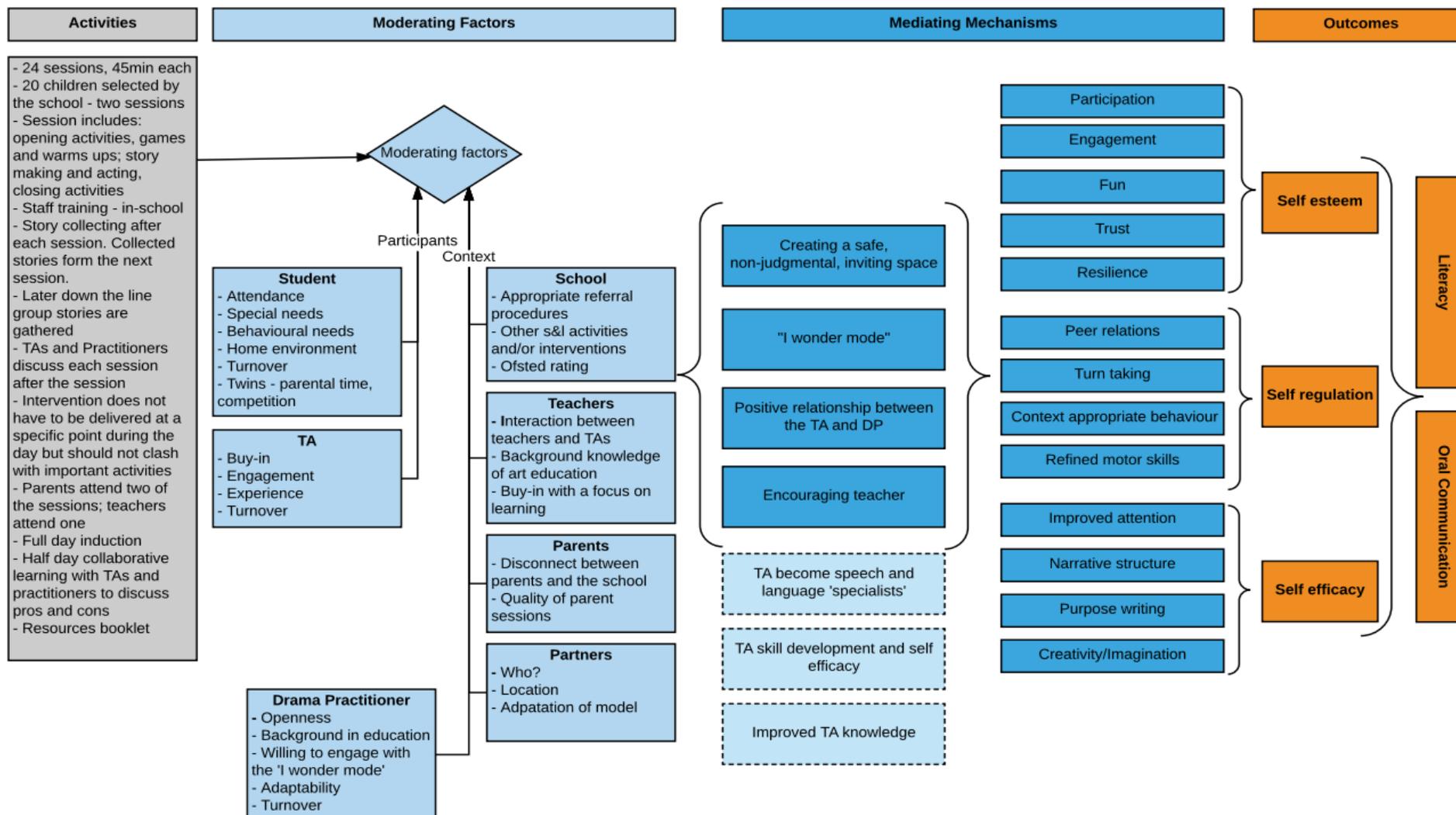
<sup>7</sup> The Communication Trust. *CPD Online Short Course: An introduction to speech, language and communication* [online]. Available at: <https://www.mendeley.com/guides/harvard-citation-guide> (accessed: 30 November 2020).

- **Decoration:** as simple as possible to reduce distractions and prevent over-stimulation;
  - **Display:** a simple display of the SB narrative cards and faces to help support visual learners.
- 9 When and how much (dosage).** SB sessions last 45 minutes and are delivered weekly during school hours, for 24 weeks, over three terms (see *Activities* in Figure 1).
- 10 Tailoring.** The intervention is intentionally designed to be adapted to each pupil's and each group's needs and interests in each session (see *Moderating factors* in Figure 1). To support certain children's (e.g., those with EAL) participation in the sessions, additional visual props, technology (such as an iPad), or support people are used to prompt engagement. The structure and delivery style of sessions is also modified to accommodate children's needs and interests.
- 11 Modifications.** There were no planned modifications to the intervention.
- 12 How well (planned).** To maintain or improve fidelity, teaching assistants (TAs) are encouraged to calm anxious pupils; develop routines so that pupils can anticipate the sessions; wear badges to indicate to pupils that there will be a session on the day; and bring certain pupils into the session space earlier to allow them time to settle in (see *Moderating factors* and *Mediating mechanisms* in Figure 1).

## Logic model

The intervention's core activity consists of the 24 weekly SB sessions (see Logic model in Figure 1). Sessions are enabled and supported by the SB resource pack; trained DPs and TAs; the induction day and CPD learning sessions with DPs, TAs and teachers; and training sessions for school staff. The SB programme ultimately aims to improve pupils' oral communication, literacy, creative self-efficacy and social skills (Outcomes). The programme aims to achieve this by first providing pupils with a safe, non-judgemental and creative environment (Mediating mechanisms). Within such an environment, pupils practise turn-taking and other context-appropriate behaviour (Mediating mechanisms), which should lead to improved social skills, such as self-regulation. Additionally, pupils practise creating narrative structure (Mediating mechanisms), which should lead to improved oral communication, literacy and creative self-efficacy. These outcomes are partly dependent on several Moderating factors, including pupil-related factors, such as session attendance, special and behavioural needs, and the everyday home environment; DP-related factors, such as their adaptability and background in education; TA-related factors, such as their buy-in and interaction with teachers; and school-related factors, such as the appropriateness of participant referrals into the program. (See Perceived Mechanisms and Outcomes in the Implementation and process evaluation (IPE) results section for a discussion of the updated logic model based on the current evaluation results.

Figure 1: Logic model



## Evaluation challenges

The attrition rate was 18.4% and 19.4%, respectively, for PIRA and RBS, the two primary outcomes. This was due to pupils changing schools (7.4% and 7.5%), being absent on testing dates (4.4% and 4.3%), and refusing testing (0.4% and 0.7%); in addition, one school did not respond to arrange testing (3.9% for both outcomes). (See Table 8 in *Attrition* for a full breakdown of these numbers)<sup>8</sup>

As indicated above, attrition due to student mobility was a particular challenge for this evaluation, and was also noted as a barrier in a separate case study (Barnes, 2018). The evaluation did impose some constraints on how attrition from the programme was handled. The SB model is best delivered with groups of ten pupils; however, upon the return from summer break, at least two schools discovered that two to three pupils in the treatment group had left the school. This posed a challenge to implementation. According to the developers, it is more difficult to run the programme with fidelity with smaller groups of pupils (eight or fewer). In a business-as-usual scenario, teachers would have filled the empty placements with other eligible pupils, who comprised the control group in the evaluation. In these schools, the teachers filled the empty positions with other pupils who met the eligibility criteria but were not part of the evaluation. It may be that these additional pupils were qualitatively different from those taking part in the evaluation, but it was agreed that this was the most practical way to maintain implementation fidelity where attrition was making it difficult to deliver the programme.

The other main challenge to the evaluation was the administration of the outcome measures. First, as the PIRA is a commonly used assessment, we did learn that at least one school had already administered the version of the assessment we were using, potentially introducing a practice effect.<sup>9</sup> However, this would not introduce any bias, as the exposure would be equal among treatment and control pupils within a given school. Second, as pupils needed to mostly complete the assessment independently, pupils with lower literacy skills struggled to engage with the test and some were not able to complete it.<sup>10</sup> Finally, the PIRA is an assessment that pupils are able to identify as an assessment, which meant some pupils were more resistant to engaging with it. The test administrators did not experience this same resistance with the RBS test, as the story-telling approach did not feel obviously like a test. The scripts provided to test administrators intentionally did not use the words 'test' and 'assessment', so as to not induce anxiety among pupils. Again, we have no reason to expect this challenge introduced any bias, given pupil-level randomisation.

## Evaluation objectives

The primary objective of this evaluation is to estimate the effect of participating in SB over the course of one school year on pupils' reading and oral communication skills. The specific research questions addressed by each element of the evaluation are further outlined below.

### Impact evaluation

#### *Primary research questions<sup>11</sup>*

- Does the programme improve reading attainment for pupils in Years 1 and 2 over the course of one year?
- Does the programme improve pupils' narrative oral skills?

---

<sup>8</sup> We note that this was in line with our assumption of a 20% attrition rate at the protocol stage (see Table 8).

<sup>9</sup> One school had flagged this to us, but we did not systematically assess this for all schools, and therefore cannot comment on the extent of this.

<sup>10</sup> The PIRA manual states that for the Year 1 Spring tests, only the phonics questions are read aloud. If pupils may be confused, administrators are able to further explain the test's requirements, but are not to read aloud any questions (unless instructed, as per above) or help with any individual words. See p. 14 in *McCarty, C. & K. Ruttle. 2018. Progress in Reading Assessment. Manual (Stage 1). Second Edition. Hodder Education.*

<sup>11</sup> Speech Bubbles Evaluation Protocol; Speech Bubbles Statistical Analysis Plan (SAP)

*Secondary research questions*

- 1 Does the programme improve pupils' social skills?
- 2 Does the programme improve pupils' perception of their ability to generate and use ideas in their work?

**Implementation and process evaluation (IPE)**

The IPE was designed to explore overarching implementation questions across all five Learning about Culture trials, as well as research questions (RQs) specific to SB. The four overarching questions were written based on cross-project similarities; however, not all questions apply to each programme due to variations in training and delivery. The relevant overarching implementation questions that are explored across all projects are detailed below. Where there were deviations from the protocol, they are noted in relation to the question and described further below.

- RQ1** In what ways was the programme implemented? What are the barriers and facilitators of delivery (Fidelity)?  
In particular:
- a Senior Leadership Team (SLT) buy-in;
  - b Delivery of training and resources: (i) the extent to which it is consistent across sites [*not answerable*]; and (ii) whether it appears to be effective in ensuring that teaching staff understand the aims and main features of the intervention;
  - c Delivery of the intervention: (i) consistent across sites [*not answerable*]; and (ii) whether it appears to facilitate children's engagement.
- RQ2** To what extent did the schools engage with the intervention in line with the intervention aims? (Responsiveness)
- RQ3** How was the quality of the intervention perceived by teachers, senior leaders and TAs? (Quality)
- RQ4** [*Changed from protocol*] How was the knowledge of the arts practitioners delivering the intervention integrated with the knowledge of teaching staff (i.e., TAs) involved? (Implementer support system) The teachers were not involved in the delivery of the intervention, instead the research question has been answered by focusing on the integration of the TAs' knowledge of the pupils.

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

- RQ5** What are the mechanisms that are taking place in the intervention and to what extent are they bringing about change? (Mechanisms)
- RQ6** [*Not answerable*] Delivered by seven separate partners across the country – to what extent is consistency ensured or the programme adapted? (Fidelity)
- RQ7** To what extent is the programme adapted by drama practitioners (DPs) and schools? (Adaption)
- RQ8** What other support do the pupils access to support their communication in both control and treatment?
- a [*Not answerable*] Are pupils in control and treatment similar? (Programme differentiation)
- RQ9** To what extent does the intervention affect the targeted children's classroom engagement and learning, particularly around engagement and communication?
- RQ10** To what extent does the intervention affect the TA, their role in school and with the class teachers? (Implementer characteristics and context)
- RQ11** To what extent do school facilities affect the intervention? (Implementation environment)

In summary, one research question (RQ6 Fidelity) and three sub-research questions (RQ1b, RQ1c, RQ8a) could not be answered. Regarding RQ1b and RQ1c (Fidelity), case study data did not produce detailed insights about the consistency of training and intervention delivery across the six schools. Regarding RQ6 (Fidelity), determining the extent of consistency across the delivery partners was not possible given that there were only six case studies. Lastly, regarding RQ8a (Programme differentiation), case study data was not able to provide detailed insights into whether

pupils in control and treatment groups are similar; this information was examined as part of the impact evaluation (see the *Pupil and school characteristics* section of this report).

## Ethics and trial registration

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

Schools were informed about the trial through initial information from the developer and formally committed to participation by signing a memorandum of understanding (MoU). A template version of the school information sheet and MoU is included in Appendix G.

This trial protocol was pre-registered at [www.controlled-trials.com](http://www.controlled-trials.com), and assigned an International Standard Randomised Controlled Trial Number (ISRCTN) of 14448319.

## Data protection

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

BIT used Article 6(1)f of the GDPR as the lawful basis for processing personal data as part of this project. This is generally known as the 'legitimate interests' basis.<sup>12</sup> BIT carried out a 'legitimate interests assessment' in support of this. The use of pupils' and teachers' personal data as part of this research was to understand the benefits to pupils of participating in this programme in terms of their academic attainment and other related benefits. This has public benefits that BIT believed are significant in terms of understanding whether this programme had the potential to benefit children in schools across England. Without processing this data, it would not have been possible to provide this quality of new evidence.

UCL used Article 6(1)e of the GDPR as the lawful basis for processing personal data as part of this project. This is generally known as the 'public task' basis. UCL has reviewed current ICO guidance<sup>13</sup> and determined that this research forms part of its performance of a task in the public interest, as one of its core purposes provided for in its Charter and Statutes. This use of data has been allocated the following UCL Data Protection Registration Number: Z6364106/2017/11/69 social research.

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

We informed pupils' parents of the proposed data processing and provided the opportunity to object to this (see Appendix H for copies of the parent information letters and data privacy notices). If parents objected, then the pupils' data were never passed on to us by schools. If a parent chose to withdraw their child's data at a later stage, then it was destroyed. The data controllers were named in the privacy information provided as part of this project and contact

---

<sup>12</sup>[https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr\\*/lawful-basis-for-processing/legitimate-interests/](https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr*/lawful-basis-for-processing/legitimate-interests/)

<sup>13</sup>ICO. *Public Task* [online]. Available at: <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/public-task/> (accessed: 30 November 2020).

details provided should they have any queries about the data we hold about them, including provision and deletion of their data.

The information provided to parents explained in clear and plain language the lawful bases for processing (although we kept the use of technical terms in the interests of keeping the language simple), the purpose of processing the data, that they could object and this would be respected, contact details of the organisation, and categories of data that we would be processing.

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

Data will be shared with the Department for Education (DfE, part of the UK Government), the Education Endowment Foundation (who funded the trial and are controllers of the EEF Data Archive), the FFT (the managers of the EEF Data Archive) and (in a form that is truly anonymised) the UK Data Archive. Details of this sharing were included in relevant Data Privacy Notices (see Appendix H).

## Project team

The project team comprised Adam Annand and Amelia Bird of London Bubble. The intervention was delivered by DPs trained by the London Bubble team.

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

## Methods

### Trial design

Table 2: Trial design

Trial design, including number of arms		Two-arm, individually randomised
Unit of randomisation		Pupil
Stratification variable(s) (if applicable)		School, year group
Primary outcomes	Variable	Reading attainment Oral communication
	Measure (instrument, scale, source)	Progress in Reading Assessment (PIRA), score range 0–25 <sup>14</sup> Renfrew Bus Story (RBS), lower score bound of 0
Secondary outcomes	Variable(s)	Social skills Creative self-efficacy
	Measure(s) (instrument, scale, source)	Social skills improvement system (SSIS): social skills sub-measure, 46 items each scored 0–3, total raw score range 0–138 Writing self-efficacy measure (WSEM): ideation sub-measure (3 questions), 3-point Likert scale, score range 3–9
Baseline for primary outcomes	Variable	Planned to be: <ul style="list-style-type: none"> <li>• Baseline reading and oral communication attainment</li> <li>• Baseline oral communication</li> <li>• Protocol deviation: FSM status, EAL status</li> </ul>
	Measure (instrument, scale, source)	Planned to be: Early Years Foundation Stage Profile (EYFSP) aggregate score (range 4–12) for four learning goals: <ul style="list-style-type: none"> <li>• understanding (FSP_COM_G02)</li> <li>• speaking (FSP_COM_G03)</li> <li>• reading (FSP_LIT_G09)</li> <li>• writing (FSP_LIT_G10)</li> </ul> EYFSP aggregate score (range 3–9) for three learning goals: <ul style="list-style-type: none"> <li>• understanding (FSP_COM_G02)</li> <li>• speaking (FSP_COM_G03)</li> <li>• listening and attention (FSP_LIT_G01)</li> </ul> <b>Protocol deviation:</b>

<sup>14</sup> This was incorrectly reported in the SAP as a range of 0–50.

		0/1 indicator of current FSM eligibility status, 0/1 indicator of EAL status. (All derived from school reports collected ahead of randomisation.)
Baseline for secondary outcomes	Variable	<p>Planned to be:</p> <ul style="list-style-type: none"> <li>• Baseline social skills</li> <li>• Baseline creative self-efficacy</li> <li>• Protocol deviation: FSM status, EAL status</li> </ul>
	Measure (instrument, scale, source)	<p>Planned to be:</p> <p>EYFSP aggregate score (range 3–9) for three learning goals:</p> <ul style="list-style-type: none"> <li>• self-confidence and awareness (FSP_PSE_G06)</li> <li>• managing feelings and behaviour (FSP_PSE_G07)</li> <li>• making relationships (FSP_PSE_G08)</li> </ul> <p>EYFSP aggregate score (range 2–6) for two learning goals:</p> <ul style="list-style-type: none"> <li>• exploring and using media and materials (FSP_EXP_G16)</li> <li>• being imaginative (FSP_EXP_G17)</li> </ul> <p><b>Protocol deviation:</b></p> <p>0/1 indicator of current FSM eligibility status, 0/1 indicator of EAL status. (All derived from school reports collected ahead of randomisation).</p>

The trial was a two-arm individually RCT testing the efficacy of the SB programme on reading attainment, oral communication, social skills and creative self-efficacy. The trial aimed to recruit at least 800 pupils across 20 schools (ultimately, 1006 pupils across 26 schools were recruited), with 40 pupils in each school randomly allocated to either the treatment arm (who received the programme) or the control group in a 50 : 50 ratio (roughly 20 in each arm). Pupils in the control group continued on a ‘business-as-usual’ basis and were not offered any incentive or other inducement to participate.

As the risk of spill-over between pupils in different arms was judged to be low, individual randomisation was considered to be acceptable for this trial. The risk was considered low because the programme required regular participation from pupils in SB sessions in order to have an impact. As teachers were not part of the intervention, there was no concern that they might change their practices. There was an acknowledged possibility that TAs who supported the SB sessions might learn new techniques or practices, which could be carried over to the general classroom setting. However, as described in the Intervention section, SB is an approach that relies upon using a safe and consistent space and a distinct set of activities over 24 regularly scheduled sessions with the leadership of a trained practitioner, which would be very difficult to replicate in a regular classroom setting. Any practices or techniques learnt by TAs that could be used in a regular classroom setting would represent a small fraction of the SB model.

The primary outcome measures were reading attainment and oral communication skills (specifically narrative recall), with secondary measures being social skills and creative self-efficacy. There were two primary outcomes because oral communication was seen as an important primary outcome alongside the more standard reading attainment indicator. For baseline measures, it was expected that we would obtain information on pupils’ performance in the EYFSP. However, due to data access issues (see Background), this was substituted by eligibility for free school meals (FSM) and whether the child has EAL, in order to improve statistical power over using an empty model.

## Participant selection

1006 pupils across 26 schools were recruited for the study.

In order to participate in the study, schools needed to:

- be located in North West England, South London or 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)<sup>15</sup>;
- have discussed participation with SB and signed an MoU detailing the conditions of participation (opt-out process, pupil data provision, end-line assessment, participation in IPE activities etc.);
- be able to refer 40 children into the study (see below for referral details).

Geographical areas of high disadvantage were targeted for recruitment. Schools with an average or above average share (14.1%<sup>16</sup>) of FSM children received priority in recruitment.

To enter the study, pupils needed to be:

- in Year 1 or 2 at the time of intervention (2018/2019)
- referred by their teachers, on the basis of guidance from the SB programme which targets the programme at children who<sup>17</sup>:
  - lack confidence in communicating;
  - have difficulty organising their thoughts and communicating them;
  - have poor attention and poor listening;
  - have not been opted-out of the study by their parents.

The process by which children entered into the study (or were randomised for those whose parents opted out<sup>18</sup>) was as follows:

- 1 Teachers referred 40 pupils across Reception and Year 1 (who will be in Year 1 or 2 at the time of the intervention).
- 2 Teachers distributed opt-out forms to the parents of the 40 children.
- 3 Once the necessary time had elapsed for opt-out return, teachers uploaded 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, English as additional language (EAL) status, teacher name and class ID.
  - b For the children whose parents opted them out, the school only supplied their first and last name, alongside teacher name and class ID, in order for us to randomise them to a group, but they were not part of the data collection. No other details were requested, and collected details were destroyed after randomisation. Pupils were randomised regardless of opt-out status, in order to ensure that all pupils had equal access to the programme. That is, opting out of the evaluation did not automatically opt the pupil out of possibly receiving the programme.
- 4 BIT randomised the 40 pupils into trial arm conditions as per the *Randomisation* section.

---

<sup>15</sup> With the exception of one pre-agreed school.

<sup>16</sup> Department of Education. (2017) Schools, pupils and their characteristics: January 2017 [online]. Available at: [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/650547/SFR28\\_2017\\_Main\\_Text.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/650547/SFR28_2017_Main_Text.pdf) (accessed: 30 November 2020).

<sup>17</sup> See Supplemental Appendices for Speech Bubbles referral guidance.

<sup>18</sup> Children who were randomised into the intervention group but were opted out of the evaluation were still able to participate in the programme.

## Outcome measures

### Baseline measures

Baseline measures for this research were planned to be drawn from the Department for Education's National Pupil Database (NPD). All participating schools were asked to provide personal information about participating pupils that would allow a reliable link to be achieved, based on current guidance from the DfE and balancing this against personal data minimisation requirements set out in data protection legislation. Using this link, it was expected that we would obtain information on pupils' performance on the Early Years Foundation Stage Profile (EYFSP). We expected to use two communications (FSP\_COM\_G02, FSP\_COM\_G03) and two literacy sub-scores (FSP\_LIT\_G09, FSP\_LIT\_G10) as baselines for the reading attainment (PIRA) outcome. We intended to use the same communication sub-scores and the listening and attention subscores (FSP\_LIT\_G01) as baselines for the oral communication (RBS) primary outcome. We had also selected three social skills sub-scores (FSP\_PSE\_G06, FSP\_PSE\_G07, FSP\_PSE\_G08) and two creativity sub-scores (FSP\_EXP\_G16, FSP\_EXP\_G17) as baselines for the social skills and creative self-efficacy outcomes, respectively.

Due to the data access issues described above, an alternative approach was taken, with its design informed by an intention to maximise the statistical power of our analysis by increasing the precision of our treatment estimates, given the data available. As such, instead of including the planned baseline measures in the model, we substituted the available demographic information that was collected about pupils ahead of randomisation (initially intended for the purposes of stratification as part of the randomisation process): specifically, eligibility for free school meals (FSM and whether the child has English as an additional language (EAL)). We include these in the model as predictors themselves. FSM and EAL status are both known to predict academic attainment and, as such, we expected this to improve power compared to an empty model.

Nevertheless, the improvement in statistical power is still likely to have fallen short of what we would have expected from including a prior attainment measure, as was planned. It is important to understand the implications of this change. First, it is important to stress that there are no expected implications for bias in our impact estimates of not having our planned baseline measures – the unbiasedness of RCT estimates derives from the randomisation, not from statistically controlling for differences at baseline. Indeed, in principle, there is no need to include any baseline measures at all in the analysis, to achieve an unbiased estimate from an RCT. Inclusion of inappropriate covariates in our analysis would have the potential to introduce bias – such inappropriate covariates are ones that could have been affected by the treatment, which is why we are including pupil characteristics from prior to randomisation. The main implication of this change is a reduction in statistical precision (i.e., the uncertainty around estimates that is inherent in all evaluations is likely larger in this evaluation than it would have been), which is manifested as wider confidence intervals (CIs) (or, equivalently, less likely to be statistically significantly different from zero for a given size of impact estimate). As with any study, there is some uncertainty around the impact estimate and there is an increased risk that headline positive or negative effect is due to this uncertainty, rather than representing a true effect as would have been the case in the presence of more explanatory power from baseline measures (and vice versa). As a result, we particularly stress the importance of statistical significance as a check on interpretation of the results in this report.

### Primary outcomes

#### Reading attainment

A core focus of the intervention is storytelling and working with narrative structure, activities that may improve pupils' reading skills (see *Mediating mechanisms* and *Outcomes* in Figure 1). This was measured using the Progress in Reading Assessment (PIRA) by Rising Stars.<sup>19</sup> PIRA is a standardised assessment of pupils' reading attainment and

profile of reading skills. It measures reading ability in the following areas: phonics, literal comprehension and reading for meaning.

The PIRA is a standardised and well-known test, which has been used in a number of prior EEF evaluations.<sup>20 21</sup> It has been shown to have high test reliability (Cronbach's alpha above 0.9), face validity (it is written to follow the national curriculum guidelines) and concurrent validity (showing a strong relationship with national test scores).<sup>22</sup> Another strength of PIRA is that tests are produced at a variety of difficulty levels, graduated by school term (e.g., 'Spring Year 1' and 'Summer Year 1'). Different versions of the PIRA test were used for the Year 1 and Year 2 cohorts. As the intervention is targeted at children with speech and language difficulties, we used a test one stage back from that which would normally be used.<sup>23</sup> As we were administering the assessments in the summer term, we used the corresponding spring versions of the test for each year group.

End-line PIRA assessments were conducted during May–June 2019 by trained research assistants (RAs) who were blind to trial arm assignment. Assessment marking was conducted by Alpha Plus, a marking service routinely used by the assessment publisher Rising Stars, and was done while blinded to trial arm assignment. Raw scores range from 0–25. As different versions of the PIRA test were used for the Year 1 and Year 2 cohorts, raw scores were standardised to have a mean of zero and a standard deviation of one prior to combining cohorts for the purpose of analysis, as specified in the SB SAP.

### **Oral communication (narrative recall skills)**

Storytelling and working with narrative structure may also improve pupils' oral communication (see *Mediating mechanisms* and *Outcomes* in Figure 1). This was measured using the Renfrew Bus Story test (RBS).<sup>24</sup> RBS is a short, standardised test that assesses narrative aspects of oral language. The administrator tells the pupil a story using a short picture book with no text as an aid. The pupil is then asked to retell the story using the picture book. 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, RBS remains the most commonly used measure (Dockerell, 2001). This assessment has some evidence of moderate test–retest reliability and high inter-rater reliability on two of the three constructs measured.<sup>25</sup>

RBS is scored along three dimensions: information (number of story details conveyed in the right order), sentence length (mean number of words in the longest five sentences), and the number of subordinate clauses. RBS raw scores are bounded at zero but do not have an upper limit. The RBS manual suggests that for pupils aged 6–7 years, the information score typically ranges from 29–35, the sentence length is 9–13 words, and the number of subordinate clauses is 3–5.

---

<sup>20</sup> McNally, S. (2016). Evaluation Protocol: An Evaluation of Teaching Assistant-Based Small Group Support for Literacy. London, United Kingdom: Education Endowment Foundation. Available at:

[https://v1.educationendowmentfoundation.org.uk/uploads/pdf/Digital\\_-\\_Small\\_Group\\_Support\\_for\\_Literacy.pdf](https://v1.educationendowmentfoundation.org.uk/uploads/pdf/Digital_-_Small_Group_Support_for_Literacy.pdf). 13 (Accessed: 30 November 2020).

<sup>21</sup> McNally, S., Ruiz-Valenzuela, J., & Rolfe, H. (2016). ABRA: Online Reading Support. London, United Kingdom: Education Endowment Foundation. Available at:

[https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation\\_Reports/EEF\\_Project\\_Report\\_A\\_BRA.pdf](https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/EEF_Project_Report_A_BRA.pdf) (Accessed: 30 November 2020).

<sup>22</sup> McCarty, C. & K. Ruttle. 2018. Progress in Reading Assessment. Manual (Stage 1). Second Edition. Hodder Education.

<sup>23</sup> This is in line with recommendations in the PIRA manual. See p.12 in McCarty, C. & K. Ruttle. 2018. *Progress in Reading Assessment*. Manual (Stage 1). Second Edition. Hodder Education.

<sup>24</sup> This measure is not available publicly for commercial reasons; see <http://www.talkingpoint.org.uk/slts/assessment-children-slc/expressive-language-assessments>

<sup>25</sup> Education Endowment Foundation. Early Years Measures Database. Available at: <https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluating-projects/early-years-measures-database/early-years-measures-database/bus-story/> (accessed 30 November 2020).

There were initially some concerns about whether children with speech delays or challenges or who had EAL would be adequately able to engage with the assessment, so BIT conducted a pilot in the year prior to the evaluation in three schools with 88 children – most of whom were currently in the SB programme. In advance of the pilot, we liaised with early year practitioners to establish an administration process that would support RAs in engaging the children to complete the assessment. The majority of children displayed full to partial engagement, with only two children not engaging at all. The results of this pilot provided confidence that the test was suitable to deliver to the vast majority of pupils taking part in the evaluation (see Appendix I for further details). The pilot also focused on the scoring of the assessment, and BIT contracted a child psychologist experienced with the assessment to advise. During this piloting, there were difficulties with establishing a high inter-rater reliability, and as such BIT opted to engage Elklan Training, a speech and language training company with expertise in administering and marking RBS, to score the assessments.

The assessment was conducted on a one-to-one basis by RAs recruited and coordinated by BIT and trained in the assessment by the same child psychologist who advised on the pilot. These RAs audio recorded the assessment, and the recordings were later transcribed. Elklan conducted scoring of the assessments using the transcriptions and audio files. The RAs, transcriptionists and Elklan staff were all blind to trial arm assignment.

## Secondary outcomes

### Social skills

The intervention involves turn-taking and other social interaction which may improve pupils' self-regulation and related social skills (see *Mediating mechanisms* and *Outcomes* in Figure 1). We assessed this for all pupils in the sample using the Social skills sub-scale of the social skills improvement system (SSIS).<sup>26</sup> The Social skills sub-scale assesses pupils' skills across the following seven sub-scales: communication, cooperation, assertion, responsibility, empathy, engagement and self-control. SSIS is a commonly used social skills assessment for young children: it is standardised, and has been used in previous EEF evaluations.<sup>27</sup> We chose to use SSIS over an equally popular instrument, the Strengths and Difficulties Questionnaire (SDQ), because the SSIS is more thorough and in-depth than the SDQ across all of the seven domains listed above. We also felt that the SSIS was a better tool to assess the types of behaviour we would expect to change as a result of participating in the programme (e.g., communication, engagement), whereas the SDQ was more oriented to identifying problematic behaviours (e.g., conduct problems, peer relationship problems).

The SSIS is designed to allow for triangulation with versions of the scale for parents and teachers to complete, as well as pupils. Due to the age of the pupils and logistical challenges of collecting data from parents, the SSIS was only collected from teachers. Reliability measures are high across all versions, with median Cronbach's alpha values in the mid to upper 0.90s for the social skills sub-scale (Gresham & Elliott, 2011). Test–retest reliability is also high for the teacher version of the scale (0.81).

The questionnaires were delivered to teachers electronically. As with all measures of social skills at this age, these had to be completed by the child's teacher. The sub-scale contains 46 items, on which teachers rate the frequency with which they observe the pupil demonstrating the behaviour; the frequency rating is then translated into point scores (Never = 0, Seldom = 1, Often = 2, Always = 3). Aggregate raw scores range from 0–138.

---

<sup>26</sup> The SSIS includes the Social skills, Problem behaviours, and Academic competence sub-scales. It is not available publicly for commercial reasons; see <https://www.pearsonclinical.com/education/products/100000322/social-skills-improvement-system-ssisrating-scales.html>

<sup>27</sup> Centre for Effective Education, Queen's University Belfast. (2016). Evaluation Protocol: Zippy's Friends. London, United Kingdom: Education Endowment Foundation. Available at: [https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation\\_Protocols/EEF\\_Project\\_Protocol\\_Character\\_Zippys\\_Friends\\_protocol.pdf](https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Protocols/EEF_Project_Protocol_Character_Zippys_Friends_protocol.pdf) (accessed 30 November 2020).

## Creative self-efficacy

As highlighted in the logic model, the impact of the intervention on literacy may have an effect through pupils' engagement with and motivation for writing (see *Mediating mechanisms* in Figure 1). For this reason, we considered pupils' self-perception of ability to generate and use ideas in their schoolwork (i.e., creative self-efficacy) as a secondary outcome measure. To measure this for all pupils in the sample, we used an adapted version of the ideation sub-measure of the writing self-efficacy measure (WSEM) 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). These adaptations were based on consultation with experts in primary literacy pedagogy at UCL.

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 also examined this sub-scale as part of the wider measure of writing self-efficacy. This measure was captured using three, three-category Likert scale items (hence, abbreviated to WSEM3; see Appendix G for the adapted scale). Each item was scored with 1–3 points. Aggregate raw scores range from 3–9. The items were asked by RAs who were blinded to trial arm assignment, after completion of the PIRA assessment.

## Sample size

Sample size calculations were used to determine the total number of schools and pupils necessary to detect the minimum effect size of interest (i.e., the minimum detectable effect size, or MDES). They were 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 are randomly allocated to either the treatment or the control group.
- **Number of treatments:** There are two trial arms (treatment and control) with 40 children in each school split equally into control and treatment groups.
- **Hypotheses:**
  - *Null hypothesis:* There is no difference in standardised PIRA or RBS scores between children who participate in the SB drama intervention and those who do not.
  - *Alternative hypothesis:* There is a difference in standardised PIRA or RBS scores between children who participate in the SB drama intervention and those who do not.
- **Attrition:** We assumed a 20% attrition rate for the end-line outcome measure for various reasons (e.g., attrition due to changing school, prolonged absence, inability to engage with the end-line assessments).<sup>28</sup> This reduced the minimum number of children per arm within the school for the purposes of sample size calculations to 17.
- **Alpha and power:** We assumed 80% statistical power and 5% significance level at the trial protocol and 2.5% at randomisation and analysis. This is because, at the time of drafting the trial protocol, it was not yet confirmed that the RBS would be used as a primary outcome. This was due to concerns over whether pupils targeted for participation in the trial would be able to engage with an RA when the assessment was administered. A pilot was conducted in June 2018 and confirmed that pupils were sufficiently engaged with the RA conducting the assessment and it was therefore decided to include the Renfrew Bus Story (RBS) as a

---

<sup>28</sup> This assumption was based on guidance published by the EEF available at the time of writing the protocol; this guidance is no longer available.

primary outcome measure.<sup>29</sup> Based on EEF statistical analysis guidelines, when using dual primary measures we applied a Bonferroni correction which reduces the alpha to 2.5%.<sup>30</sup>

- Test–retest correlation:** The baseline achievement measure that was planned to be used is the Early Years Foundation Stage Profile (EYFSP). The only estimate for the test–retest correlation between the EYFSP and the reading assessment PIRA was 0.61. This was based on an unpublished analysis from the Fisher Family Trust (FFT), conducted at the end of Year 1 for a prior EEF trial (ABRA: Online Reading Support).<sup>31</sup> However, given that our study targets a specific population, we opted for a conservative estimate of 0.3 for the PIRA. For the RBS assessment of oral communication skills, we did not have any information on its correlation with EYFSP and therefore, at the protocol stage, we conducted power calculations for a range of values between 0–0.8. At the randomisation stage, we assumed a test–retest correlation of 0.3. At the analysis stage, as discussed above, we were unable to access the EYFSP baseline measures as planned. We have substituted these measures with FSM eligibility (in the past year) and whether the child has EAL status, but have conservatively assumed zero impact of these variables on statistical power.
- Minimum detectable effect size (MDES):** 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. The required MDES for the full sample analysis was 0.20 SD (Cohen’s *d*).
- MDES for free school meals (FSM) sub-group:** Although the trial was not powered to detect an effect on the FSM sub-group as the primary population of interest, we specified an FSM sub-group analysis in the protocol and therefore estimated the MDES for this sample. We assumed that the FSM sub-group is 14.1% of the total sample (based on data from DfE statistics),<sup>32</sup> and maintained the expected test–retest correlation coefficient value of 0.30. However, as previously explained, we were not able to access the NPD EVERFSM6 variable and analyses, so instead we used data provided by the school which indicated whether the pupil was eligible for FSM at point of enrolment (which was the year prior to implementation).

Sample size calculations at the protocol, randomisation and analysis stages are summarised in Table 7 (see *Impact evaluation results*). At the protocol stage, calculations using the above assumptions indicated that, in total, 960 pupils across 23 schools were required to achieve an MDES of 0.2 SD (assuming the use of only the PIRA outcome at the time of protocol, and therefore no Bonferroni correction, as discussed above). It was assumed that 130 pupils were FSM, indicating an MDES of 0.53 SD<sup>33</sup> for this sub-group. Based on these calculations, a recruitment target of 25 schools was set to provide a buffer against attrition.

At randomisation, 100634 pupils across 26 schools were included in the sample. This indicated an MDES of 0.21 SD<sup>35</sup> for PIRA and RBS, a slight increase which was due to the use of a Bonferroni correction to adjust for two primary outcomes confirmed at this stage (all other assumptions were the same as at the protocol stage). Within this sample, 344 pupils were eligible for FSM at the time of recruitment, based on information provided by the schools.

---

<sup>29</sup> Although the delivery of the assessment did not prove challenging, the marking did. We sought help from speech and language specialists, who helped resolve marking inconsistencies and agreed to be involved in RA training prior to end-line data collection.

<sup>30</sup> Statistical analysis guidance for EEF evaluations (March 2018). Retrieved from: [https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing\\_a\\_Protocol\\_or\\_SAP/EEF\\_statistical\\_analysis\\_guidance\\_2018.pdf](https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Protocol_or_SAP/EEF_statistical_analysis_guidance_2018.pdf)

<sup>31</sup> [https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation\\_Reports/EEF\\_Project\\_Report\\_ABRA.pdf](https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/EEF_Project_Report_ABRA.pdf)

<sup>32</sup> [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/650547/SFR28\\_2017\\_Main\\_Text.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/650547/SFR28_2017_Main_Text.pdf)

<sup>33</sup> In the SAP, this is incorrectly reported as 0.51 SD, which does not correctly reflect the stated assumed pupil-level attrition rate of 20%.

<sup>34</sup> 1009 pupils were originally randomised, but at the point of outcome data collection schools informed us that three pupils had been opted out of the evaluation, so we have excluded them from the sample entirely.

<sup>35</sup> In the SAP, this is incorrectly reported as 0.17 SD, which does not correctly reflect the stated assumed pupil-level attrition rate of 20% and the assumed Bonferroni correction to adjust for two primary outcomes.

This indicated an MDES of 0.32 SD<sup>36</sup> for PIRA and RBS for the FSM sub-group. This decrease in MDES reflects a much larger proportion of FSM pupils than assumed (32.9% and 35.5% in the intervention and control groups, respectively, compared to the 14.1% assumed at the protocol stage).

At the time of analysis, the sample size was 821 pupils across 25 schools for the PIRA, and 811 pupils across 25 schools for the RBS. This indicated an MDES of 0.22 SD for the PIRA and RBS outcomes, which is similar to that estimated at the protocol and randomisation stages. Within this sample, 273 pupils for the PIRA outcome were eligible for FSM (276 pupils for RBS), which indicated an MDES of 0.34 SD for the PIRA and RBS outcomes. This slight increase in MDES for the overall and FSM samples is explained partly by the inability to access baseline measures as planned (discussed above) and partly by the lower number of pupils at the analysis stage. The overall sample sizes reflect attrition rates of 18.4% for PIRA and 19.4% for RBS, which are close to the 20% assumed at the protocol stage (see Table 8 for a summary of attrition).

## Randomisation

Pupils were referred by teachers using the eligibility criteria described previously. Pupils were randomised into trial arms within schools following the completion of the opt-out process and data transfer. Randomisation was conducted by BIT staff using the data analysis and statistical software Stata (see Appendix K for the code used to carry out the randomisation, including the stable seed to allow for replication of the process). School and year level (Years 1 and 2) were the stratification variables. The randomisation proceeded in the following steps for each school:

1. If more than 40 pupils were referred,<sup>37</sup> we contacted the schools and asked them to restrict the sample to 40 pupils.<sup>38</sup>
2. Pupils in the school were stratified by year level into two blocks (each with 20 pupils).
3. Within each stratum, a random number was generated for each pupil and half the pupils within each stratum were assigned to the treatment group.

We progressively checked that the resulting control and treatment groups were balanced in terms of the absolute number of children who would not participate in data collection due to their parents opting them out of the study. This was to ensure the number of children allocated to each trial arm did not become unduly unequal. Analysis was not undertaken while blinded to randomisation. To minimise bias, the analysis code was quality assured by an independent researcher at BIT and a member of the UCL research team.

## Statistical analysis

### Primary analysis

All analyses were performed using R. The evaluation had two primary outcomes: reading attainment, as measured by PIRA, and oral communication, as measured by RBS. As different versions of the PIRA test were used for the Year 1 and Year 2 cohorts, raw scores were standardised to have a mean of zero and a standard deviation of one prior to combining cohorts for the purpose of analysis. For the RBS, raw scores were analysed.

---

<sup>36</sup> In the SAP, this is incorrectly reported as 0.29 SD, which does not correctly reflect the stated assumed pupil-level attrition rate of 20% and the assumed Bonferroni correction to adjust for two primary outcomes.

<sup>37</sup> Schools were discouraged from doing this and asked to prioritise referring those students they believed would most benefit from the intervention. The targeting of these students is how the intervention is used more generally.

<sup>38</sup> This process deviated from the trial protocol, which stated that '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.'

As we tested two primary outcome measures, we applied a Bonferroni correction (in line with EEF 2018 statistical guidance),<sup>39</sup> and thus reported with 97.5% CI. We estimated impacts as intention to treat (ITT) effects.

In the evaluation protocol and SAP, we stated our intent to carry out the analyses using the following ordinary least squares (OLS) model in order to estimate the ITT impact of the intervention (the same analysis model was used for both primary outcomes):

$$Y_i = \alpha + \beta_1 Treat_i + \beta_2 X_i + \beta_3 Year_i + \beta_4 School_i + \epsilon_i,$$

where

$Y_i$  is the relevant primary outcome measure (PIRA or RBS score) for student  $i$

- $Treat_i$  is a binary indicator for the treatment assignment for student  $i$  (1 if the student is assigned to treatment; 0 if not);
- $X_i$  is a vector of the relevant baseline attainment measured through aggregated EYFSP learning goal scores for student  $i$ ;
- $Year_i$  is a binary variable for the year group (1 for Year 2 and 0 for Year 1);
- $School_i$  is a vector of school fixed effects for  $n-1$  schools;
- $\epsilon_i$  is the individual heteroskedasticity-robust error term.

However, due to data access issues discussed above, we are unable to estimate this model due to the unavailability of the EYFSP scores as  $X_i$ . Instead, we estimate the following model in which  $X_i$  has been replaced with FSM eligibility and EAL status (as discussed above):

$$Y_i = \alpha + \beta_1 Treat_i + \beta_2 FSM_i + \beta_3 EAL_i + \beta_4 Year_i + \beta_5 School_i + \epsilon_i,$$

where everything is as per the planned model, except that  $FSM_i$  is whether individual  $i$  was eligible for free school meals in the previous school year and, similarly,  $EAL_i$  is whether individual  $i$  is recorded as having English as an additional language.

Note that while this model is a deviation from the evaluation protocol and statistical analysis plan, it was planned and reported to the EEF ahead of analysis being carried out. This model has not been altered depending on the significance of any variables included (i.e., no variables were removed due to being statistically insignificant).

## Secondary analysis

The secondary outcomes were creative self-efficacy, as measured by the WSEM3, and social skills, as measured by SSIS. Analyses were carried out on raw scores, using the revised primary analysis model specified above, except replacing  $Y_i$  with the WSEM3 or SSIS scores. As discussed above, this model is a deviation from protocol and was made due to data access problems rendering the baseline EYFSP measures unavailable.

## Analysis in the presence of non-compliance

As described in the trial protocol, pupil-level compliance in this trial was defined as having attended at least 16 of the 24 SB sessions (66%) prior to the collection of outcome data. This definition was agreed in collaboration with the developer. Attendance for each pupil was recorded by the drama practitioner (DP) on a roster provided by the delivery team and provided to the evaluation team at the end of the year. We estimated treatment effects for all four outcome measures for compliers using a complier average causal effect (CACE) analysis. The CACE estimation used a two-stage least squares (2SLS) approach, which is based on the revised primary ITT model described above. At the first stage, we estimated a model of compliance:

---

<sup>39</sup>[https://educationendowmentfoundation.org.uk/public/files/Grantee\\_guide\\_and\\_EEF\\_policies/Evaluation/Writing\\_a\\_Protocol\\_or\\_SAP/EEF\\_statistical\\_analysis\\_guidance\\_2018.pdf](https://educationendowmentfoundation.org.uk/public/files/Grantee_guide_and_EEF_policies/Evaluation/Writing_a_Protocol_or_SAP/EEF_statistical_analysis_guidance_2018.pdf)

$$Comply_i = \alpha + \beta_1 Treat_i + \beta_2 FSM_i + \beta_3 EAL_i + \beta_4 Year_i + \beta_5 School_i + \mu_i$$

where

$Comply_i$  is a binary indicator for whether student  $i$  met the minimal compliance threshold

- $Treat_i$  is a binary indicator for the treatment assignment for student  $i$
- $\mu_i$  is the individual heteroskedasticity-robust error term;
- all other variables are as defined in the revised primary analysis ITT model above.

At the second stage, we used the predicted values from  $Comply_i$  to estimate a model of the outcome measure  $Y_i$ :

$$Y_i = \alpha + \beta_1 \widehat{Comply}_i + \beta_2 FSM_i + \beta_3 EAL_i + \beta_4 Year_i + \beta_5 School_i + \varepsilon_i,$$

where

- $Y_i$  is the relevant primary or secondary outcome measure for student  $i$  (specified in the *Outcome measures* section)
- $\widehat{Comply}_i$  are the predicted levels of compliance with the programme from the first equation;
- $\varepsilon_i$  is the individual heteroskedasticity-robust error term;
- All other variables are as defined in the revised primary analysis ITT model above.

We conducted this analysis using the 'ivreg' function from the AER package in R. We note the deviation to protocol, with these models being based on the revised primary outcome ITT model, rather than the planned primary outcome ITT model, which is for the same underlying reasons of data access described above.

## Missing data analysis

We describe and summarise the extent of missing data in the primary outcomes, and in the model associated with the analysis. We also describe reasons for missing data, including withdrawal of the school from the study, a pupil leaving the school, a pupil refusing testing, and a pupil being absent on the day(s) of data collection.

In the SAP, we outlined a missing data strategy for the primary analysis. For missing outcome data, the strategy notes that a complete case analysis will be run. For missing predictor data, this strategy notes that multiple imputation (MI) would be conducted if more than 5% of data in the model is missing. We note that there was no missing predictor data in this evaluation, so this part of the strategy was never implemented. We outline it below for transparency.

In the case where more than 5% of predictor data is missing, we planned to first consider whether the missing data was missing at random (MAR). We would do this by using logistic regression to test whether the missing status can be predicted from all variables in the revised analysis model. Where predictability was confirmed (i.e., if the estimated coefficient on any of the explanatory variables in the model was significantly different from zero at the 5 percent significance level), we planned to estimate a MI model using a fully conditional specification. MI was planned to be carried out using the Markov Chain Monte Carlo (MCMC) method to predict the missing values prior to the analysis of treatment effects. We would then estimate the treatment effect using the imputed data in the primary analysis model and compare our result with the primary analysis (conducted on complete cases only). MI was planned to be implemented using the 'mice' package in R.

Analysis using the multiply imputed data set was planned to be used as a sensitivity analysis (i.e., we would base confirmation of the effectiveness of the treatment on complete case analysis only, but would assess the sensitivity of the estimate to missingness using the estimates from the multiply imputed data set). If the complete case analysis model implied effectiveness but the imputed estimate did not, we would assume that the missing data is missing not at random to such an extent as to invalidate our conclusion of effectiveness.

As an additional check (not specified in the SAP), we tested whether missing outcome data may be considered missing completely at random. We ran a logistic regression model to predict missingness in primary outcome data (PIRA and RBS score) using all variables in the primary analysis model:

$$MissingStatus_i \sim \text{bernoulli}(p_i)$$

$$\text{logit}(p_i) = \alpha + \beta_1 Treat_i + \beta_2 FSM + \beta_3 EAL_i + \beta_4 Year_i + \beta_5 School_i,$$

where  $MissingStatus_i$  is a binary indicator for whether the primary outcome (PIRA or RBS score) is missing for student  $i$ . All other variables are the same as in the primary analysis model. If any of the variables (other than school fixed effects) are statistically significant predictors of missingness (at the 0.05 significance level), this would suggest that missing outcome data are not missing completely at random, and instead are missing at random conditional on these variables.

### Sub-group analyses

We conducted analysis on the primary and secondary outcomes for the sub-group of FSM pupils. We originally planned to do this using the EVERFSM\_6\_P variable from the NPD, which indicates pupils who have ever in the past six years been registered for FSM. We planned to modify the original primary and secondary analysis models, with the addition of an interaction between treatment assignment and FSM status, to assess whether there is a significant difference in the treatment effect between FSM students and others:

$$Y_i = \alpha + \beta_1 Treat_i + \beta_2 FSMEver_i + \beta_3 Treat_i \times FSMEver_i + \beta_4 X_i + \beta_5 Year_i + \beta_6 School_i + \epsilon_i,$$

where

- $Y_i$  is the relevant primary or secondary outcome measure (specified in the *Outcome measures* section) for student  $i$
- $Treat_i$  is a binary indicator for the treatment assignment for student  $i$  (1 if the student is assigned to treatment; 0 if not)
- $FSMEver_i$  is a binary indicator for student  $i$ 's EVERFSM\_6\_P status (1 if the student has been recorded as eligible for FSM; 0 if not)

All other variables are as specified in the original primary and secondary analysis models.

As discussed above, the inability to access data required us to estimate the following revised model:

$$Y_i = \alpha + \beta_1 Treat_i + \beta_2 FSM_i + \beta_3 Treat_i \times FSM_i + \beta_4 EAL_i + \beta_5 Year_i + \beta_6 School_i + \epsilon_i,$$

where  $FSM_i$  is whether individual  $i$  was eligible for FSM in the previous school year (information provided by the schools). This model is identical to the revised primary analysis ITT model, with the addition of an interaction between  $Treat_i$  and  $FSM_i$ . One difference between the planned EVERFSM\_6\_P variable and the FSM variable that we are now using is that the latter only reflects FSM eligibility in the past year, rather than the current year or any historical eligibility in the past six years. Given that pupils are only in Year 1 and Year 2, the lack of information on historical FSM eligibility should have minimal impact on the evaluation.

Additionally, in line with EEF guidelines, we estimated a separate model on the restricted sample of FSM pupils using the revised primary and secondary analysis models specified above.

### Additional analyses and robustness checks

No additional statistical analyses were planned. However, we conducted several additional exploratory analyses.

#### *EAL sub-group analysis*

We considered whether there is evidence of differential treatment effects between EAL and non-EAL pupils. This is relevant because the pupil eligibility criteria for this trial included whether pupils lack confidence in communicating, have difficulty in communicating thoughts, or have poor listening, as identified by teachers. The identification of these features may be associated with whether a pupil is EAL. In line with this, over half of the sample at randomisation consisted of EAL pupils (56.2% and 56.9% of the intervention and control groups, respectively). Similarly, to the FSM

sub-group analysis, we added to the primary and secondary analysis models an interaction term between the treatment variable and membership of the EAL sub-group:

$$Y_i = \alpha + \beta_1 Treat_i + \beta_2 FSM_i + \beta_3 EAL_i + \beta_4 EAL_i \times Treat_i + \beta_5 Year_i + \beta_6 School_i + \epsilon_i.$$

In addition, we estimated separate models on the restricted samples of EAL and non-EAL pupils using the revised primary and secondary analysis models specified above.

#### *Randomisation inference*

We carried out alternative statistical inference for the primary and secondary analyses, using randomisation inference to provide useful information on the extent to which there is variation between these different approaches to statistical inference. Randomisation inference is a method of conducting statistical inference using the uncertainty inherent in the randomisation process regarding the assignment of units in the trial to the treatment arm, rather than any appeal to an external sample and sampling variation (i.e., focusing on internal rather than external validity, see Cunningham, 2021). We used 5000 permutations to generate the null distribution. The analysis was done using the ‘ri2’ package in R. As this was not planned in the SAP, this should be considered exploratory and will not be used to guide interpretation of the main results.

#### *Clustered standard error robustness check*

Although this is a pupil-randomised trial, pupils are nevertheless naturally clustered within schools. This was confirmed by the observation of a non-zero intra-cluster correlation (ICC) for each outcome (see section below on Estimation of ICC). Such clustering can inflate standard error estimates and increase the risk of false positives in statistical testing. As such, we re-ran the primary and secondary analyses using standard errors clustered at the school level. Given the small number of clusters (25 schools at analysis), the use of traditional cluster-robust standard errors is inappropriate. We therefore used the wild cluster bootstrapping approach to generate *t*-statistics as implemented in the ‘clusterSEs’ R package. We generated 5000 bootstrap permutations.

#### *Estimation of effect sizes*

Hedges’ *g* effect size (Hedges, 1981) was calculated as follows:

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

where our conditional estimate  $\underline{x}_1 - \underline{x}_2$  was recovered from  $\beta_1$  in the primary ITT analysis model.

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

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

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

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

$$J(n_1 + n_2 + 2) = \frac{\Gamma((n_1 + n_2 + 2)/2)}{(\sqrt{((n_1 + n_2 + 2)/2)}) \Gamma((n_1 + n_2 + 2 - 1)/2)}.$$

If this proved intractable, we used the following approximation instead:

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

97.5% confidence intervals of the effect size (for primary analyses which required Bonferroni corrections; 95% confidence intervals for all other analyses) were estimated by inputting the upper and lower confidence limits from the regression model into the effect size formula.

#### **Estimation of ICC**

We estimated the ICC of the primary and secondary outcome measures at the school level by estimating an empty variance components model, as follows:

$$Y_{ij} = \alpha + \gamma_j + \epsilon_{ij},$$

where

- $Y_{ij}$  is the relevant primary or secondary outcome measure for pupil  $i$  in school  $j$ ;
- $\gamma_j$  is the school-level random effect;
- $\epsilon_{ij}$  is the individual error term.

The school-level random effect was assumed to be normally distributed and uncorrelated with the individual-level errors. An empty variance components model is used to facilitate comparability between trials (and in line with EEF guidance). The ICC itself was estimated from this model using the following equation:

$$\rho = \text{var}(\gamma_i) / (\text{var}(\gamma_i) + \text{var}(\epsilon_i)).$$

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

#### *Longitudinal analysis*

No longitudinal analyses were planned but we propose that, given the paucity of evidence around arts / creativity, the EEF specifically commits to following up this trial cohort to understand if there are long-term differences between the intervention and control arms of the study.

## Implementation and process evaluation (IPE)

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

### **IPE aims and approach**

The purpose of the IPE was to understand how the SB intervention was delivered within schools and, in particular, to better understand the barriers and facilitators to implementation and delivering the intervention with fidelity. The IPE was also used to understand how the SB intervention compared to the existing support available to pupils with speech and language difficulties, and to determine the cost to schools of delivering the intervention. Multiple sources of data were triangulated to address the IPE research questions. The primary approach to IPE data collection consisted of collecting case study data from six schools that were participating in this study. Each case study involved the following methods:

- a semi-structured interview with:
  - the teaching assistant (TA) that co-delivered the intervention;
  - the drama practitioner (DP) that co-delivered the intervention;
  - a Year 1 or Year 2 teacher;
  - a member of the SLT;
- an observation of a SB session;
- informal discussions with participating children;
- where possible, a joint interview was also conducted with the TA and DP following the observed SB session.

In addition, a survey was sent to all participating schools, administrative data were collected at induction (September 2018), and mid-point training sessions (January 2019) were observed. Intervention manuals and guidance were also reviewed to inform topic guides for interviews and data analysis. The research questions and the data collection methods used to address them are shown in Table 3.

Table 3: IPE methods overview

Data collection method	Data analysis method	Type of school	Participant groups / activity type	Target number of participants	Achieved number of participants / activities	Total achieved number of participants / activities	Research questions addressed
Semi-structured Interviews	Framework approach	Case study schools	Senior leadership team (SLT)	1 per case study	1 per case study	6	RQ1 (Fidelity), RQ3 (Quality), RQ5 (Mechanisms), RQ8 (Programme differentiation), RQ9 (Reach), RQ10 (Implementer characteristics), RQ11 (Implementation environment)
			Drama practitioner (DP)	1 per case study	1 per case study	6	RQ1 (Fidelity), RQ4 (Implementer support system), RQ5 (Mechanisms), RQ7 (Adaptation), RQ10 (Implementer characteristics), RQ11 (Implementation environment)
			Teaching assistant (TA)	1 per case study	1 per case study	6	RQ1 (Fidelity), RQ3 (Quality), RQ4 (Implementer support system), RQ5 (Mechanisms), RQ7 (Adaptation), RQ8 (Programme differentiation), RQ9 (Reach), RQ10 (Implementer characteristics), RQ11 (Implementation environment)
			Teacher	1 per case study	1 per case study	6	RQ1 (Fidelity), RQ3 (Quality), RQ5 (Mechanisms), RQ8 (Programme differentiation), RQ9 (Reach), RQ10 (Implementer characteristics)
			Pupils	1 per case study	1 per case study	6	RQ5 (Mechanisms)
Informal group interview							
Observation			Speech Bubbles (SB) session	1 per case study	1 per case study	6	RQ1 (Fidelity), RQ5 (Mechanisms), RQ7 (Adaptation), RQ11 (Implementation environment)
Surveys	Descriptive statistics	All participating schools	Teaching assistants (TAs)	26	N/A	21	RQ1 (Fidelity), RQ3 (Quality), RQ7 (Adaptation), RQ8 (Programme differentiation), RQ9 (Reach), RQ10 (Implementer characteristics), RQ11 (Implementation environment)
Cost interview	Framework approach	4 cost case study schools	Teachers	4	1 per case study	4	Cost

# Methods

## Sampling and recruitment

### Case studies

Six case study schools were selected, using a purposive sampling approach to capture the range of schools participating in the SB intervention during the 2018/2019 academic year. The primary sampling criteria were:

- 1 level of engagement in training (defined as high where the school's TA had attended two CPD sessions, and low where the TA had only attended one CPD session at the point of sampling);
- 2 the proportion of pupils on the school roll receiving FSM (defined as high where the percentage was 23% or higher, the median for all SB schools, and low where the percentage was below 23%, using information obtained from UK Government, 2019).

The secondary sampling criteria were:

- 1 range of delivery partners (DPs from eight theatres were involved in delivering the SB intervention);
- 2 the proportion of pupils participating in SB with EAL (defined as high where over 67.5% of the cohort had EAL, medium when between 32.5% and 67.5% had EAL, and low where less than 32.5% of the cohort had EAL);
- 3 Ofsted rating (recorded as Outstanding, Good or Requires improvement, using information obtained from UK Government, 2019).

Table 4 sets out characteristics for the six case study schools that were recruited. Information is not provided on the theatre delivery partners to preserve anonymity, but all of the schools sampled had a DP employed by a different theatre.

Table 4: Characteristics of case study schools

Characteristics	Number
TA engagement in training	
High (2 sessions attended)	3
Low (1 session attended)	3
Free school meals (FSM) (whole-school data)	
Low (22% or lower)	3
High (23% or higher)	3
English as an additional language (EAL) (SB cohort)	
30–40%	2
41–50%	0
51–60%	0
61–70%	1
71–80%	0

81–90%	1
91–100%	2
Ofsted rating	
Outstanding	2
Good	2
Requires Improvement	1
No rating as academy converter school (school is now rated 'good')	1

### *Surveys (all participating schools)*

Two surveys (baseline and follow-up at the end of the intervention period) were developed and administered to SB co-facilitators (i.e., TAs) in all schools taking part in the trial. Unfortunately, the baseline survey had a low response rate (54%, n=14), so it did not produce meaningful data. Instead, the questions were incorporated into the follow-up survey (see Appendix P). The follow-up survey had a 81% response rate (n=21). The follow-up survey indicated that 81% of the respondents were TAs (n=17), while some of the other SB co-facilitators had other roles within the schools such as being a learning assistant (see answers to question 1.2 in Appendix P for the full range of roles). As the surveys were a census and not a sample, CI are not given for response frequencies; in addition, as the schools taking part in the trial were not a representative sample of all primary schools in the UK, it is not appropriate to generalise the findings from the surveys beyond this group of schools.

## **Data collection**

### *Case studies*

Sampled schools were contacted by email, and where schools agreed to take part, a date was arranged for a researcher to visit. The researchers did not inform London Bubble about which schools they intended to visit. All visits took place between March and June 2019. At the visits, the following data were collected: a semi-structured, audio-recorded interview with (i) the SB DP, (ii) the TA involved in co-delivering the intervention, (iii) a teacher, whose pupils were participating in the intervention, and (iv) a member of the school's SLT; an observation of a SB session; and informal discussions with participating children. It was also set out in the protocol that a joint interview would be conducted with the TA and DP after the observation. This was done wherever possible (at three out of six case study schools), but was not always feasible due to the time available in the school day.

The interviews were conducted using structured guides that focused on exploring the following:

- usual practice in terms of supporting children with speech, language and communication difficulties;
- the context in which the intervention was implemented;
- the facilitators and barriers to implementation;
- pupils' engagement in intervention;
- the perceived impact of the intervention;
- the mechanisms underlying this change.

Interviewees were informed that their participation in the interview would remain anonymous, that they could withdraw at any time and that they did not have to answer any questions they did not want to. Discussions with children covered the content of SB sessions, what they thought of the sessions and how SB compared to other lessons. The researcher recorded notes during and immediately after each discussion. Full interview guides can be found in Appendix N.

Observations focused on capturing detail about the space in which the session was delivered; the content of the session; the perceived confidence of the DP and TA co-delivering the session and how they shared roles; and children's perceived engagement with the session. Observation notes were recorded first in field notes, then transferred to on a structured pro forma (which can be found in the Supplemental Appendices). The observations proforma was developed collaboratively in partnership with the UCL team, who brought their respective subject matter expertise. Specifically, Dominic Wyse has writing, music and mixed-methods research expertise; Gemma Moss has expertise in early literacy development; and Andrew Burn has expertise in English, media and drama. Strategies to reduce bias were implemented, including having two researchers (e.g., a BIT and a UCL team member) conducting the initial observations, discussing and agreeing upon the final observation notes that were recorded; one of those researchers continued to conduct the subsequent observations in the other case study schools. Further, the observation data were used to help researchers probe effectively during the interviews and to deepen understanding of observed practice.

#### *Surveys (all participating schools)*

The research questions and programme logic model were used to inform the design of the baseline survey. Data from interviews, as well as feedback from the delivery team, were used to inform the design of the follow-up surveys. The survey questions focused on TAs' self-reported confidence to co-deliver the intervention; facilitators and barriers to attending training and delivering SB sessions; views on working with the SB drama specialist; and the perceived impact of the intervention for participating children and the TAs themselves. There was also one question about the usual support that is available to children with speech, language and communication difficulties. Full survey details can be found in Appendix P.

For both the baseline and follow-up survey, the key contact at each school was emailed a link to the online survey platform SmartSurvey. The email stated that the survey should be completed by the TA who had been involved in co-delivering the SB intervention. For both surveys, reminders were sent to schools who did not initially complete the survey. For the follow-up survey, a researcher attended a training event as a prompt to TAs to complete it. Schools who did not complete the follow-up survey were also contacted by phone and given the opportunity to complete the survey over the phone. All data was collected by the end of July 2019.

#### *Administrative data*

For the purposes of case study sampling, data were collated in February 2019 on the number of CPD sessions attended by the TA; the name of each school's drama specialist and the theatre they were employed by; and the proportion of EAL children participating in SB at each school. As a measure of engagement with the intervention, data were also collected at the end of the trial on the number of SB sessions (out of a total of 24 possible sessions) attended by each pupil taking part.

### **Data analysis**

#### *Case studies*

Verbatim transcripts of the interviews and notes from the observations were analysed using the Framework approach (Ritchie et al., 2014). Firstly, emerging themes were identified through familiarisation with the data by reviewing the observation notes and transcripts. The analytical framework was then created using a series of matrices in Excel, each relating to an emergent theme. The columns in each matrix represented the key sub-themes drawn from the findings and the rows represented individual participants interviewed or schools observed.

The interview and observation data were then summarised in the appropriate cell, which meant that all data relevant to a particular theme were noted, ordered and accessible, facilitating a systematic approach to analysis that was grounded in participants' and schools' accounts. Analysis involved working through the charted data to draw out the range of schools' experiences and participants' views, as well as identifying similarities, differences and links between them. Thematic analysis (undertaken by looking down the theme-based columns in the Framework) identified concepts and themes, and the case-based analysis (undertaken by comparing and contrasting rows in the Framework) enabled links within cases to be established and cases compared and contrasted with each other. During the analytical process a balance was maintained between deduction (using existing knowledge and the research questions to guide the analysis) and induction (allowing concepts and ways of interpreting experience to emerge from

the data). The *Implementation and process evaluation (IPE) results* section is organised based on the identified themes and sub-themes, which are outlined and described in their respective subsections.

Multiple strategies were employed by the researchers to increase the credibility (i.e., accurate representation of the data), transferability (i.e., potential to apply the findings to other settings), dependability (i.e., traceable, logical analytical process) and confirmability (i.e., being grounded and traceable to the raw data) of the findings, with the ultimate aim of reducing bias during the analytical process (Hannes, 2011).

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

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

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

### Surveys

Due to a low response rate (54%), data were not analysed for the baseline survey. For the follow-up survey, data were cleaned by identifying schools that had returned two or more survey responses. In these instances, the most recent survey completed by a respondent who identified as having co-delivered the intervention (typically a TA), was used for analysis. Data from those co-delivering the intervention were prioritised where there were multiple responses from the same school, because it was assumed that they would have been most involved with delivery, and therefore, their views and experiences were most relevant.

Prior to cleaning the survey data set, there were 31 responses to the follow-up survey. Following cleaning, 21 responses remained (out of 26 schools taking part in the trial), giving a response rate of 81%. Stata (version 14) was used to analyse the data to obtain descriptive statistics for each question. Percentage scores are reported, where relevant, in the *Implementation and process evaluation (IPE) results* section. Complete survey findings are provided in Appendix P.

### Costs

The evaluation gathered three key categories of data: direct marginal costs (which will form the basis of the cost per pupil); pre-requisites (which will be reported separately from the cost per pupil); and school staff time. The data was gathered in two ways. Firstly, the evaluators requested information from delivery partners on how much they charged schools for delivering the intervention as part of the evaluation, and how much they will charge schools in the future excluding any funding or subsidy that is associated with delivering the intervention as part of this evaluation. The latter

data is used in calculating the cost per pupil; the former data is to ensure there is clarity about the precise nature of the data that is being requested and transparency of the approach.

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

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

The resources required to deliver the intervention were most influenced by staff time and any related marginal costs (such as materials needed for implementation and travel and subsistence). We assumed these costs were most likely to be related to the amount a school has to spend per pupil,<sup>40</sup> and the nature of the local area and school population, the most relevant indicator for which is the percentage of pupils eligible for FSM. That is, school spending on the programme is likely related to school financial resources. Therefore, for each trial we purposively selected one school in each of the following four categories (see Table 5), in order to capture the range and diversity of costs to implement the programme.

Table 5: Categories and category definitions based on which schools were selected

Category	Category definition
#1	School with percentage of FSM in top half of participating schools and spend per pupil in top half of participating schools
#2	School with percentage of FSM in top half of participating schools and spend per pupil in bottom half of participating schools
#3	School with percentage of FSM in bottom half of participating schools and spend per pupil in top half of participating schools
#4	School with percentage of FSM in bottom half of participating schools and spend per pupil in bottom half of participating schools

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

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

<sup>40</sup> Data obtained from <https://www.gov.uk/school-performance-tables>

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

## Timeline

Table 6: Timeline

Dates	Activity	Staff responsible / leading
October 2017–March 2018	School recruitment	London Bubble
January–March 2018	MoU signing and ensuring children are referred by teachers	London Bubble
April 2018–May 2018	Distribution of opt-out forms to parents of referred children	BIT
June 2018–July 2018	Rolling randomisation as data upload and opt-out process completion confirmed	BIT
September 2018–June 2019	Intervention delivery	London Bubble
September 2018	Observation of induction day / training session	BIT
September 2018	Twilight session for all Key Stage 1 (KS1) school staff	London Bubble
October 2018	IPE baseline survey	BIT
November 2018	Parent session observation	BIT
January 2019	Collaborative learning / CPD session observation	BIT
February 2019–March 2019	Sampling for case studies	BIT
March 2019–June 2019	Case study school staff interviews, discussions with pupils and observations	BIT
June 2019–July 2019	End-line assessments (PIRA, RBS test, WSEM) administered by RAs (BIT), SSIS completed by teachers Compliance data collection	BIT
July 2019	Evaluation and celebration event for teaching assistants (TAs)	London Bubble
July 2019	Follow-up survey for TAs	BIT
July–August 2019	Marking of PIRA end-line assessments Transcription and marking of Renfrew Bus Story (RBS) assessments Data entry of WSEM	Hodder, contracted by BIT Elklan, contracted by BIT BIT
July–September 2019	Collation and cleaning of outcomes and compliance data in readiness for upload to ONS SRS for linkage with DfE National Pupil Database (NPD) extract	UCL and BIT
September 2019–October 2019	Cost evaluation data collection	BIT

January–July 2020	Project paused waiting conclusion of data sharing agreement necessary for upload of project data to the ONS SRS for linkage with NPD.	
August–November 2020	Project resumes with planned revisions as a result of delays in achieving data sharing agreement. Impact analysis and report writing. BIT leads on the data analysis with agreed deviations from the published statistical analysis plan (SAP).	UCL and BIT

## Impact evaluation results

### Participant flow including losses and exclusions

The flow of participants is detailed in Figure 2 for the two primary outcomes. Of the 219 schools that were approached, 26 schools (12%) agreed to participate in the trial and met the eligibility criteria. Across these schools, 1006 pupils were randomly allocated using a stratified, individual-level randomisation, as described above, resulting in 502 pupils in the intervention group and 504 pupils in the control group. Figure 2 shows that there were some pupils who could not be reached for follow-up. For the PIRA outcome, this included 91 pupils in the intervention group, leaving 411 pupils for analysis, and 94 pupils in the control group, leaving 410 pupils for analysis. For the RBS outcome, this included 97 pupils in the intervention group, leaving 405 pupils for analysis, and 98 pupils in the control group, leaving 406 pupils for analysis. These numbers include one entire school that could not be reached for testing, meaning that the total number of schools analysed was 25.

The MDES estimated at various points of the trial is reported in Table 7. For the PIRA and RBS outcomes, this was 0.2 SD at the protocol design stage, which increased to 0.21 SD<sup>41</sup> at randomisation. This slight increase is due to the use of a Bonferroni correction to adjust for two primary outcomes, which were only confirmed at the randomisation stage.<sup>42</sup> At the analysis stage, the MDES for the PIRA and RBS outcomes increased slightly to 0.22 due to the inability to access baseline measures as planned (discussed above). We note that the MDES at the analysis stage for both outcomes is similar to that estimated at the protocol and randomisation stages.

---

<sup>41</sup> In the SAP, this is incorrectly reported as 0.17 SD, which does not correctly reflect the stated assumed pupil-level attrition rate of 20% and the assumed Bonferroni correction to adjust for two primary outcomes.

<sup>42</sup> As indicated in the SAP, at the time of drafting the trial protocol, it was not yet confirmed that the RBS outcome would be used as a primary outcome, and therefore a 5% significance level was assumed. When this was confirmed at randomisation, a 2.5% level was assumed.

Figure 2: Participant flow diagram (2 arms) for both primary outcomes (PIRA and RBS)

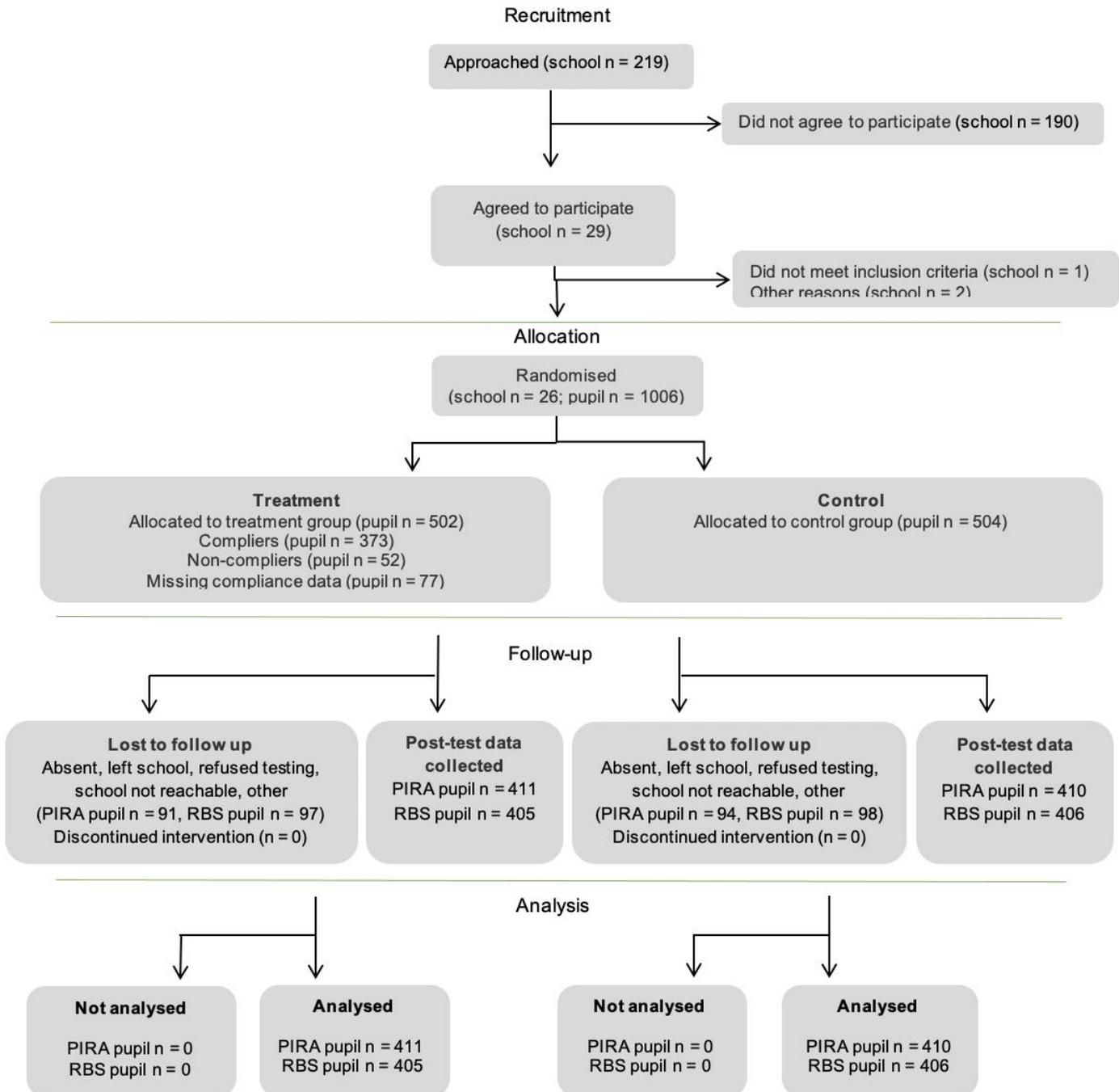


Table 7: Minimum detectable effect size (MDES) at different stages

		Stage											
		Protocol				Randomisation				Analysis			
		All pupils		FSM		All pupils		FSM		All pupils		FSM	
		PIRA	RBS	PIRA	RBS	PIRA	RBS	PIRA	RBS	PIRA	RBS	PIRA	RBS
MDES		0.2	0.2	0.53 <sup>43</sup>	0.53	0.21 <sup>44</sup>	0.21	0.32 <sup>45</sup>	0.32	0.22	0.22	0.34	0.34
Pre-/post-correlations	Level 1 (pupil)	0.3	0.3 <sup>46</sup>	0.3	0.3	0.3	0.3	0.3	0.3	0	0	0	0
Intra-cluster correlations (ICCs)	Level 3 (school)	–	–	–	–	–	–	–	–	–	–	–	–
Alpha		0.05	0.05	0.05	0.05	0.025 <sup>47</sup>	0.025	0.05 <sup>48</sup>	0.05	0.025	0.025	0.05	0.05
Power		0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
One-sided or two-sided?		two-sided	two-sided	two-sided	two-sided	two-sided	two-sided	two-sided	two-sided	two-sided	two-sided	two-sided	two-sided
Average cluster size		–	–	–	–	–	–	–	–	–	–	–	–
Number of schools	Total:	23	23	23	23	26	26	26	26	25	25	25	25
Number of pupils	Intervention	460	460	65	65	502	502	165	165	411	405	131	134
	Control	460	460	65	65	504	504	179	179	410	406	142	142

<sup>43</sup> Note that, in the SAP, the MDES of 0.51 did not correctly reflect the stated assumed pupil-level attrition rate of 20%.

<sup>44</sup> Note that, in the SAP, the MDES of 0.17 did not correctly reflect the stated assumed pupil-level attrition rate of 20% and the assumed Bonferroni correction to adjust for two primary outcomes.

<sup>45</sup> As above, this number now correctly reflects the assumptions stated in the SAP.

<sup>46</sup> At the protocol design stage, power calculations for RBS assumed a range of pre-test/post-test correlations due to a lack of prior information on the magnitude of this correlation. For simplicity, here we use the assumption of 0.3, in line with the PIRA outcome.

<sup>47</sup> As indicated in the SAP, at the time of drafting the trial protocol, it was not yet confirmed that the RBS outcome would be used as a primary outcome, and therefore a 5% significance level was assumed. When this was confirmed at randomisation, a 2.5% level was assumed.

<sup>48</sup> Note that the sample size table on page 5 in the SAP incorrectly states the alpha level as 2.5% for the FSM sub-group analysis. A Bonferroni correction was only planned to be applied for the primary analysis, as stated in the primary analysis specification section on p. 6 in the SAP.

	Total:	920	920	130	130	1006 <sup>49</sup>	1006	344	344	821	811	273	276
--	--------	-----	-----	-----	-----	--------------------	------	-----	-----	-----	-----	-----	-----

## Attrition

Attrition from randomisation to analysis for the PIRA and RBS outcomes is summarised in Table 8 for the PIRA outcome; 36 pupils in the intervention group (37 pupils for RBS) and 38 pupils in the control group (38 for RBS) changed schools before testing. Twenty-three (23) pupils in the intervention group (21 for RBS) and 21 pupils in the control group (22 for RBS) were absent on both the main and follow-up days of testing. Three pupils in the intervention group (2 for RBS) and 1 pupil in the control group (5 for RBS) refused testing. One entire school (19 pupils in the intervention group and 20 in the control group) did not respond to arrange testing. Data were not available for other reasons for a further 10 pupils in the intervention group (18 for RBS) and 14 pupils in the control group (13 for RBS).<sup>50</sup> For both the PIRA and RBS outcomes, the attrition rate was evenly distributed across intervention and control groups (for PIRA, the attrition rate was 18.1% for the intervention group and 18.7% for the control group; for RBS, it was 19.3% and 19.4%, respectively). The attrition rate overall was 18.4% for PIRA and 19.4% for RBS, which is slightly better than the rate of 20% assumed at the protocol stage.

Table 8: Pupil level attrition from the trial (primary outcome)

		Intervention		Control		Total	
		PIRA	RBS	PIRA	RBS	PIRA	RBS
Number of pupils	Randomised	502	502	504	504	1006	1006
	Analysed	411	405	410	406	821	811
Pupil attrition (from randomisation to analysis)							
Overall	Number	91	97	94	98	185	195
	Percentage	18.1%	19.3%	18.7%	19.4%	18.4%	19.4%
Changed schools	Number	36	37	38	38	74	75
	Percentage	7.2%	7.4%	7.5%	7.5%	7.4%	7.5%
Absent on day of testing	Number	23	21	21	22	44	43
	Percentage	4.6%	4.2%	4.2%	4.4%	4.4%	4.3%

<sup>49</sup> 1009 pupils were originally randomised, but at the point of outcome data collection schools informed us that three pupils had been opted out of the evaluation, so we have excluded them from the sample entirely.

<sup>50</sup> One reason is that follow-up testing was not conducted for schools with fewer than three pupils absent on the main testing days; other reasons are unknown (i.e., collected data about pupils' testing status and associated information is missing for this small minority of pupils).

Pupil refused testing	Number	3	2	1	5	4	7
	Percentage	0.6%	0.4%	0.2%	1.0%	0.4%	0.7%
School did not respond to arrange testing	Number	19	19	20	20	39	39
	Percentage	3.8%	3.8%	4.0%	4.0%	3.9%	3.9%
Other reasons	Number	10	18	14	13	24	31
	Percentage	2.0%	3.6%	2.8%	2.6%	2.4%	3.1%

Note. Percentage sums may be inexact due to rounding.

## Pupil and school characteristics

Table 9 summarises the baseline pupil-level characteristics of intervention and control pupils as randomised. As per the SAP, for all binary characteristics we report and descriptively compare counts and percentages. In general, it shows that intervention and control pupils were similar to each other. The intervention group had a slightly lower proportion of FSM pupils (32.9%) than the control group (35.5%). The two groups had a similar proportion of EAL pupils (intervention 56.2% vs. control 56.9%) and Year 1 pupils (intervention 48% vs. control 49.8%).

Both the intervention and control groups differed from the national-level figures. The groups had slightly higher proportions of FSM pupils (intervention 32.9%, control 35.5%), compared to the national average of 29.6%. They also had substantially higher proportions of EAL pupils (intervention 56.2%, control 56.9%), compared to the national average of 16.3%.

Table 9: Baseline characteristics of pupils as randomised

Pupil-level (categorical)	National-level mean (Key Stage 2) <sup>51</sup>	Intervention group		Control group	
		<i>n/N</i> (missing)	Count (%)	<i>n/N</i> (missing)	Count (%)
Previous year FSM <sup>52</sup>	29.6	502/1006 (0)	165 (32.9)	504/1006 (0)	179 (35.5)
Previous year non-FSM	70.4	502/1006 (0)	337 (67.1)	504/1006 (0)	325 (64.5)
EAL	16.3	502/1006 (0)	282 (56.2)	504/1006 (0)	287 (56.9)
Non-EAL	83.7	502/1006 (0)	220 (43.8)	504/1006 (0)	217 (43.1)

<sup>51</sup> National-level figures reflect percent FSM and EAL at Key Stage 2 (from *Compare School Performance KS2 data, 2018–2019*), given that school-level public data on KS1 FSM and EAL were not available. This comparison is limited by differences between these two stages of education.

<sup>52</sup> Due to the inability to access NPD data (discussed above), we only have data from schools on whether pupils in the trial were currently eligible for FSM at the time of recruitment, rather than EVERFSM\_6.

Year 1	–	502/1006 (0)	241 (48)	504/1006 (0)	251 (49.8)
Year 2	–	502/1006 (0)	261 (52)	504/1006 (0)	253 (50.2)

Table 10 summarises the baseline school-level characteristics and how they compare to national-level figures. Schools in the sample were all urban, compared to the 70.8% of schools at the national level. There were 12 community schools, 11 academies and three other types of school. Ten schools were rated by Ofsted as Outstanding, 13 were rated as Good and one was rated as Requires improvement.

Table 10: Baseline characteristics of schools as randomised

School-level (categorical)		National-level proportion	N (missing)	Count (%)
Location	Urban	70.8	26 (0)	26 (100)
	Rural	29.2	26 (0)	0 (0)
School type	Community	36.0	26 (0)	12 (46.2)
	Academy	32.4	26 (0)	11 (42.3)
	Other	31.6	26 (0)	3 (11.5)
Ofsted rating	Outstanding	16.6	26 (2)	10 (38.5)
	Good	68.1	26 (2)	13 (50)
	Requires improvement	10.4	26 (2)	1 (3.8)
School-level (continuous)		National-level mean <sup>53</sup>	N (missing)	Mean (SD)
Key Stage 1 (KS1) average performance		16.3	26 (4)	16.0 (1.4)
Percent free school meals (FSM) at Key Stage 2 (KS2) <sup>54</sup>		29.6	26 (4)	51.8 (15.7)

<sup>53</sup> Compare School Performance KS2 data 2018–2019.

<sup>54</sup> These figures reflect percent FSM and EAL at KS2 (from Compare School Performance KS2 data, 2018–2019), given that school-level public data on KS1 FSM and EAL were not available. This comparison is limited by differences between these two stages of education.

Percent English as an additional language (EAL) at KS2	16.3	26 (4)	59.2 (24.1)
--------------------------------------------------------	------	--------	-------------

Tables 11 and 12 present the analogous balance characteristics for the groups as analysed for the PIRA and RBS outcomes. Group differences here are overall comparable to the sample as randomised. The main difference is that in the analysed sample for the PIRA outcome, the intervention group had a slightly lower proportion of EAL pupils (56% vs. the control group (58.8%). All other comparisons between intervention and control groups, and between the groups and the national-level figures, are very similar to the sample as randomised. All analyses control for FSM, EAL and year group.

Table 11: Baseline characteristics of pupils as analysed (PIRA outcome)

Pupil-level (categorical)	National-level mean (Key Stage 2) <sup>55</sup>	Intervention group		Control group	
		n/N (missing)	Count (%)	n/N (missing)	Count (%)
Previous year free school meals (FSM) <sup>56</sup>	29.6	411/821 (91)	131 (31.9)	410/821 (94)	142 (34.6)
Previous year non-FSM	70.4	411/821 (91)	280 (68.1)	410/821 (94)	268 (65.4)
English as an additional language (EAL)	16.3	411/821 (91)	230 (56)	410/821 (94)	241 (58.8)
Non-EAL	83.7	411/821 (91)	181 (44)	410/821 (94)	169 (41.2)
Year 1	–	411/821 (91)	201 (48.9)	410/821 (94)	203 (49.5)
Year 2	–	411/821 (91)	210 (51.1)	410/821 (94)	207 (50.5)

<sup>55</sup> National-level figures reflect percent FSM and EAL at KS2 (from *Compare School Performance KS2 data, 2018–2019*), given that school-level public data on KS1 FSM and EAL were not available. This comparison is limited by differences between these two stages of education.

<sup>56</sup> Due to the inability to access NPD data (discussed above), we only have data from schools on whether pupils in the trial were currently eligible for FSM at the time of recruitment, rather than EVERFSM<sub>6</sub>.

Table 12: Baseline characteristics of pupils as analysed (RBS outcome)

Pupil-level (categorical)	National-level mean (Key Stage 2) <sup>57</sup>	Intervention group		Control group	
		<i>n/N</i> (missing)	Count (%)	<i>n/N</i> (missing)	Count (%)
Previous year free school meals (FSM) <sup>58</sup>	29.6	405/811 (97)	134 (33.1)	406/811 (98)	142 (35)
Previous year non-FSM	70.4	405/811 (97)	271 (66.9)	406/811 (98)	264 (65)
English as an additional language (EAL)	16.3	405/811 (97)	226 (55.8)	406/811 (98)	230 (56.7)
Non-EAL	83.7	405/811 (97)	179 (44.2)	406/811 (98)	176 (43.3)
Year 1	–	405/811 (97)	199 (49.1)	406/811 (98)	204 (50.2)
Year 2	–	405/811 (97)	206 (50.9)	406/811 (98)	202 (49.8)

Table 13 presents the school characteristics for the group as analysed. Given that only one school is excluded from the analysis sample, characteristics are very similar to the sample as randomised.

Table 13: Baseline characteristics of schools as analysed (PIRA and RBS outcomes)

School-level (categorical)		National-level proportion	<i>N</i> (missing)	Count (%)
Location	Urban	70.8	25 (0)	25 (100)
	Rural	29.2	25 (0)	0 (0)
School type	Community	36.0	25 (0)	12 (48)

<sup>57</sup> National-level figures reflect percent FSM and EAL at KS2 (from *Compare School Performance KS2 data, 2018-2019*), given that school-level public data on KS1 FSM and EAL were not available. This comparison is limited by differences between these two stages of education.

<sup>58</sup> Due to the inability to access NPD data (discussed above), we only have data from schools on whether pupils in the trial were currently eligible for FSM at the time of recruitment, rather than EVERFSM<sub>6</sub>.

	Academy	32.4	25 (0)	10 (40)
	Other	31.6	25 (0)	3 (12)
Ofsted rating	Outstanding	16.6	25 (2)	10 (40)
	Good	68.1	25 (2)	12 (48)
	Requires improvement	10.4	25 (2)	1 (4)
School-level (continuous)	National-level mean <sup>59</sup>		<i>N</i> (missing)	Mean (SD)
Key Stage 1 (KS1) average performance		16.3	25 (4)	16.1 (1.3)
Percent free school meals (FSM) at Key Stage 2 (KS2) <sup>60</sup>		29.6	25 (4)	51.1 (15.8)
Percent English as an additional language (EAL) at KS2		16.3	25 (4)	59.7 (24.6)

## Outcomes and analysis

### Primary analysis

The two primary outcome measures were the PIRA, a measure of reading attainment, and the RBS, a measure of oral communication. Figure 3 shows the histograms for these outcomes across all pupils. Data for both outcomes broadly follow a normal distribution, with a slight leftward skew for the PIRA. The RBS score data show a small peak at 0, reflecting a proportion of pupils that either did not speak at all during the testing or did not say anything relevant for scoring. This could slightly attenuate impact estimates, but otherwise has minimal implications for the analysis model. Across the intervention and control groups, the PIRA outcome was standardised within each year group to have a mean score of 0 and standard deviation of 1 to account for differences in the difficulty of the test versions between year groups. The RBS outcome has a raw mean score of 25.1 (SD = 11.3).

<sup>59</sup> *Compare School Performance KS2 data 2018–2019*,

<sup>60</sup> These figures reflect percent FSM and EAL at KS2 (from *Compare School Performance KS2 data, 2018-2019*), given that school-level public data on KS 1 FSM and EAL were not available. This comparison is limited by differences between these two stages of education.

Figure 3: Histograms of outcomes for primary outcomes (PIRA and RBS)

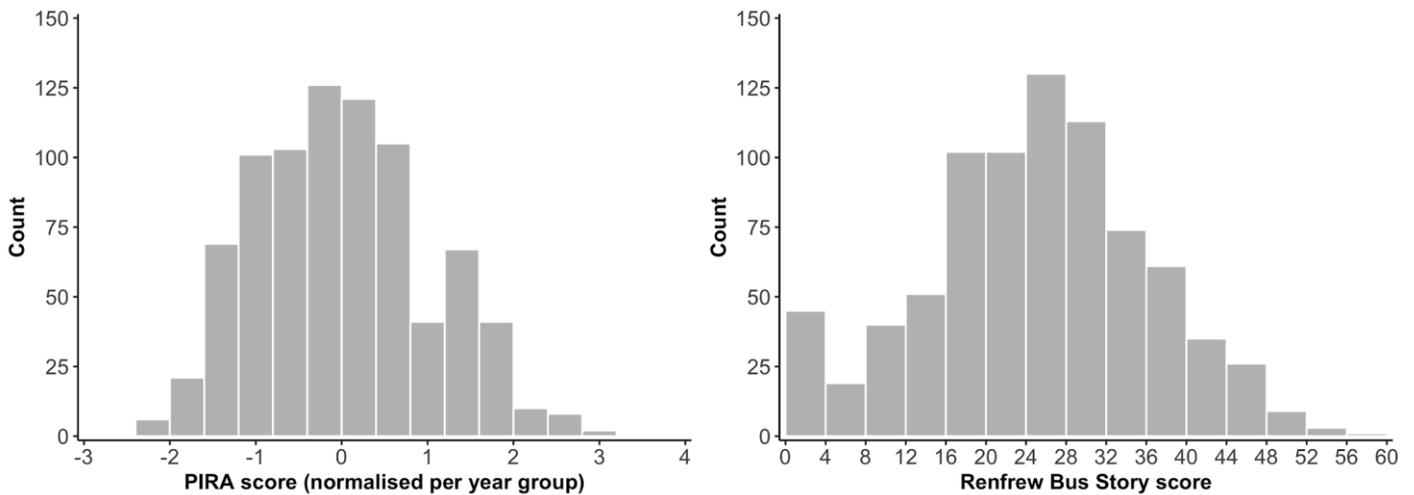


Table 14 presents the results of the analysis for the two primary outcome measures. It shows the unadjusted mean for the PIRA score in the SB intervention group (−0.03) and the unadjusted mean in the control group (0.03). After adjusting for covariates in the analysis model, the mean difference between the two groups is −0.05, which translates into a Hedges’ *g* effect size of −0.05 (a lower score in the intervention group compared to the control group), which is not statistically significant ( $p = 0.93$ ). *This means that the SB intervention did not have a statistically significant effect on the PIRA score.*

Table 14: Primary analysis

Outcome	Unadjusted means				Effect size		
	Intervention group		Control group		Total <i>n</i> (intervention; control)	Hedges’ <i>g</i> (97.5% CI)	<i>p</i> -value
<i>n</i> (missing)	Mean (95% CI)	<i>n</i> (missing)	Mean (95% CI)				
PIRA	411 (91)	−0.03 (−0.13, 0.06)	410 (94)	0.03 (−0.06, 0.13)	821 (411, 410)	−0.05 (−0.2, 0.1)	0.93
RBS	405 (97)	24.92 (23.84, 26.00)	406 (98)	25.27 (24.15, 26.39)	811 (405, 406)	−0.04 (−0.19, 0.11)	> 0.99

Note. *p*-values for the two primary outcomes are adjusted for multiple comparisons using the Bonferroni correction and 97.5% CI are reported for the treatment effect size.

For the RBS score, the unadjusted mean in the SB intervention group is 24.92 and that in the control group is 25.27. After adjusting for covariates in the analysis model, the mean difference is −0.46. This translates into a Hedges’ *g* effect size of −0.04 (a lower score in the intervention group compared to the control group), which is also not statistically significant ( $p > 0.99$ ). *This means that the SB intervention did not have a statistically significant effect on the RBS score.*

## Secondary analysis

The two secondary outcome measures were the WSEM3, a measure of creative self-efficacy, and the SSIS, a measure of social skills. Figure 4 shows the histograms for these outcomes across all pupils. Data for the WSEM3 broadly follow a normal distribution, with no implications for the analysis model. Data for the SSIS show a slight rightward skew, suggestive of a small ceiling effect (although the distribution tails off toward the top scores). This could slightly attenuate impact estimates but otherwise has minimal implications for the analysis model. Across both experimental arms, the WSEM3 has a mean score of 6.23 (SD = 1.47) and the SSIS has a mean score of 91.75 (SD = 22.45).

Figure 4: Histograms of outcomes for secondary outcomes (WSEM3 and SSIS)

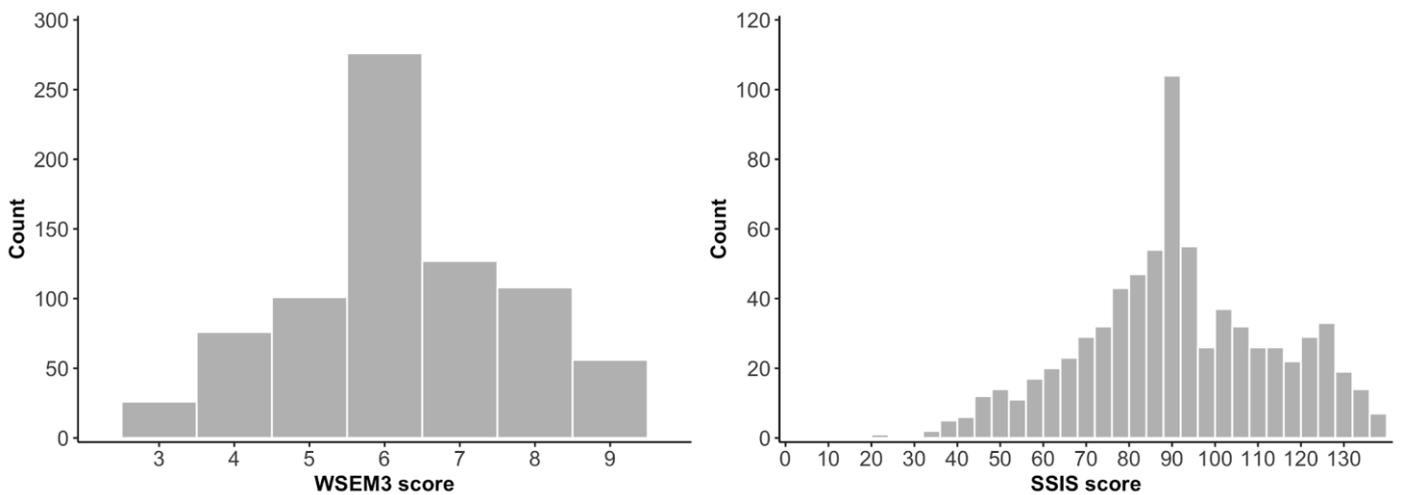


Table 15 presents the results of the analysis for the two secondary outcome measures. For the WSEM3 score, the unadjusted mean in the SB intervention group is 6.27 and that in the control group is 6.20. The Hedges'  $g$  effect size for the difference between groups is 0.05 (a higher score in the intervention group compared to the control group) but this difference is not statistically significant ( $p = 0.47$ ). *This means that the SB intervention did not have a statistically significant effect on the WSEM3 score.*

Table 15: Secondary analysis

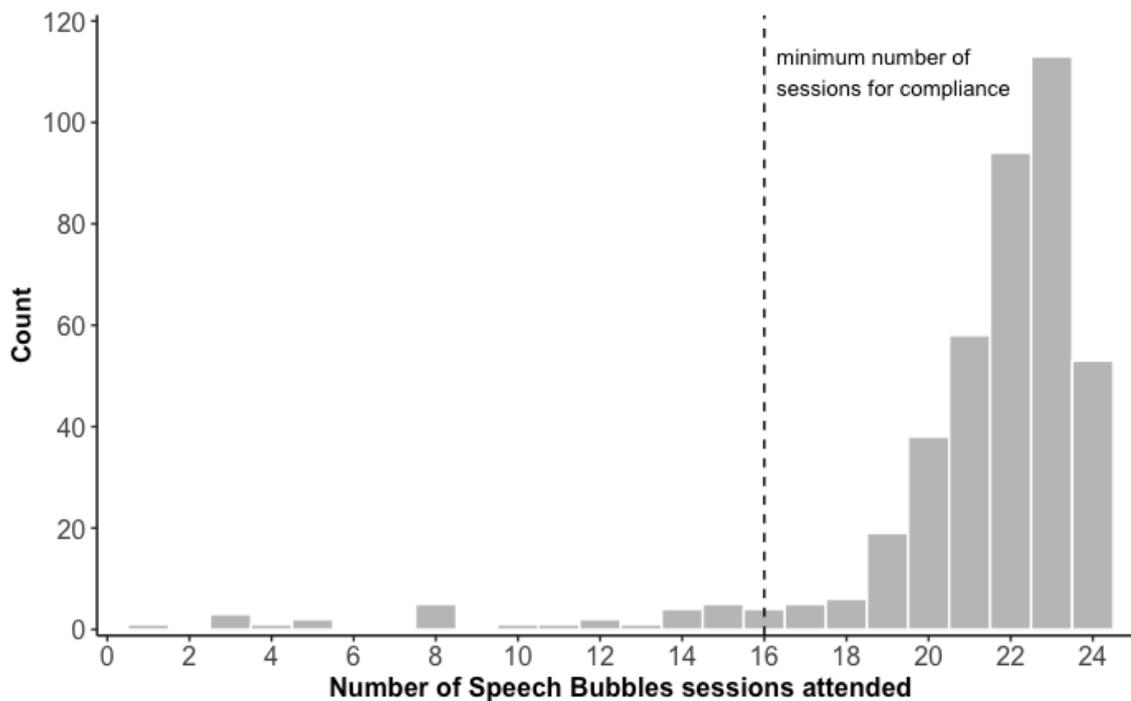
Outcome	Unadjusted means				Effect size		
	Intervention group		Control group				
	$n$ (missing)	Mean (95% CI)	$n$ (missing)	Mean (95% CI)	Total $n$ (intervention; control)	Hedges' $g$ (95% CI)	$p$ -value
WSEM3	377 (125)	6.27 (6.12, 6.41)	393 (111)	6.2 (6.05, 6.35)	770 (377, 393)	0.05 (-0.09, 0.19)	0.47
SSIS	383 (119)	92.00 (89.69, 94.31)	363 (141)	91.49 (89.25, 93.73)	746 (383, 363)	0.03 (-0.12, 0.17)	0.71

For the SSIS score, the unadjusted mean in the SB intervention group is 92.00 and that in the control group is 91.49. The Hedge's  $g$  effect size is 0.03 (a higher score in the intervention group compared to the control group), but this difference is also not statistically significant ( $p = 0.71$ ). *This means that the SB intervention did not have a statistically significant effect on the SSIS score.*

### Analysis in the presence of non-compliance

In line with the protocol, compliance in the SB intervention is defined for pupils in the intervention group as attending at least 16 out of 24 SB sessions. Weekly attendance for SB sessions was recorded by the DPs. We were not able to obtain these compliance data from three schools (58 pupils in the intervention group). This means that out of the 502 pupils in the intervention group (at randomisation), we have compliance data for 88.4% of pupils (444 pupils). Of these pupils, 87.2% were compliers, attending at least 16 sessions. Pupils attended a median of 22 sessions. Figure 5 shows a histogram of the number of sessions attended among pupils in the intervention group.

Figure 5: A histogram of the number of SB sessions attended among pupils in the intervention group



In order to examine whether pupil non-compliance may be diluting the estimate of the intervention effect, we estimate the complier average causal effect (CACE) for each of the primary and secondary outcomes. This analysis estimates the effect of the intervention among compliers.<sup>61</sup> We note that, given the high compliance rate of 87.2%, results from this analysis should be broadly similar to the ITT analyses. The analysis was conducted on pupils for whom we had data both on compliance and for each of the primary and secondary outcomes. We estimate the CACE using the 'ivreg' function from the AER package in R (see Appendix M for syntax).

A summary of the CACE results and further details is presented in Table 16. There were no statistically significant effects for any of the four outcomes. For all outcomes, the directions of effects were the same as in the primary and

<sup>61</sup> As described in the challenges section, some schools added additional pupils to the programme, as high levels of attrition made it challenging to deliver the programme with fidelity. Where pupils were added, their names were added to the register (handwritten additions). In all but one case, the added pupils were not part of the evaluation. One control group pupil was recorded as attending 22 sessions. This is accounted for in the instrumental variable CACE analysis by indicating this pupil to be a 'complier' with the intervention.

secondary analyses, and effect sizes were also broadly comparable. This suggests that there is no evidence of treatment effects for either primary or secondary outcome among pupils who complied with the intervention.

Table 16: Complier average causal effect (CACE) analysis results

Outcome	Intervention N (missing)	Control N (missing)	Weak instruments first-stage F test ( $p$ -value)	Compliance / treatment correlation ( $p$ -value)	CACE Hedges' $g$ (95% CI)	CACE $p$ -value
PIRA	363 (139)	362 (142)	F(1, 699) = 6120.32 ( $p < 0.001$ )	0.95 ( $p < 0.001$ )	-0.07 (-0.22, 0.08)	0.36
RBS	364 (138)	360 (144)	F(1, 698) = 6428.58 ( $p < 0.001$ )	0.95 ( $p < 0.001$ )	-0.04 (-0.19, 0.1)	0.57
WSEM3	334 (168)	350 (154)	F(1,658) = 6028.77 ( $p < 0.001$ )	0.95 ( $p < 0.001$ )	0.01 (-0.14, 0.17)	0.88
SSIS	346 (156)	327 (177)	F(1, 649) = 5635.17 ( $p < 0.001$ )	0.94 ( $p < 0.001$ )	0.02 (-0.14, 0.18)	0.8

### Missing data analysis

In the SAP, we outlined a missing data strategy for the primary analysis. For missing predictor data, this strategy notes that multiple imputation would be conducted if more than 5% of data in the model is missing. In the current case, there is no missing predictor data (0% for FSM status, EAL status, year group, school ID), so imputation is not necessary.

For missing outcome data, the strategy notes that a complete case analysis will be run. The extent of missing outcome data and reasons for this missingness are summarised in Table 8 (see the *Attrition* section). The extent of missing data is comparable between groups, with the intervention group missing 18.1% of data for the PIRA outcome (19.3% for the RBS outcome) and the control group missing 18.7% of data (19.4% for the RBS outcome). The proportion of data missing for known reasons (i.e., pupils changing schools, being absent on the day of testing, refusing testing, or the school not responding to arrange testing) is also comparable between groups. Data missing for unknown reasons is equally comparable between groups, and comprises just 2% of the intervention group's data and 2.8% of the control group's data for the PIRA outcome (3.6% and 2.6% for the RBS outcome, respectively).

As an additional check, we ran a logistic regression model to predict missingness in outcome data (for PIRA and RBS outcomes) using all variables in the analysis model (treatment, FSM, EAL, year group and school ID). For both outcomes, there were no significant predictors of missingness, other than fixed effects for two schools (for both,  $p < 0.05$ ; see Table L1 in Appendix L for results). Therefore, we assume that the data is missing completely at random. Moreover, because data missing for unknown reasons comprises such a small proportion of the overall data, we do not expect this missingness to affect the analysis conclusions.

### Sub-group analyses

As is standard in all EEF-funded evaluations, we tested whether there is evidence of differential effects on each of the primary and secondary outcomes between FSM and non-FSM pupils. To do this, we added to the primary and secondary analysis models an interaction term between the treatment variable and membership of the FSM sub-group. For all outcomes, the estimates for this interaction term are small and not statistically significant, providing little

evidence of a differential effect between FSM and non-FSM pupils. In addition, we used the primary and secondary analysis models to estimate treatment effects among the sub-sample of FSM pupils. For all outcomes, the treatment effects are small and not statistically significant. For the primary outcomes, the direction of the effects is the same as for the full sample, with lower scores in the intervention group compared to the control group. For the WSEM3, the direction of the effect is also the same as for the full sample, with higher scores in the intervention group compared to the control group. For the SSIS, the direction of the effect is in the opposite direction as for the full sample, with lower scores in the intervention group compared to the control group. Table 17 summarises these results.

Table 17: Free school meals (FSM) sub-group analyses

Outcome	Interaction term Hedges' <i>g</i> (95% CI)	Interaction term <i>p</i> -value	Sub-group total <i>n</i> (intervention, control)	Sub-group treatment Hedges' <i>g</i> (95% CI)	Sub-group treatment <i>p</i> - value
PIRA	-0.05 (-0.32, 0.23)	0.73	273 (131, 142)	-0.04 (-0.29, 0.2)	0.73
RBS	-0.06 (-0.32, 0.2)	0.66	276 (134, 142)	-0.08 (-0.3, 0.15)	0.50
WSEM3	-0.03 (-0.34, 0.27)	0.83	256 (118, 138)	0.08 (-0.19, 0.36)	0.56
SSIS	-0.08 (-0.39, 0.23)	0.6	260 (131, 129)	-0.05 (-0.31, 0.22)	0.73

### Additional analyses and robustness checks

As outlined in the SAP for this trial, no additional analyses were planned for this trial. However, we carried out several exploratory analyses. These were not pre-specified and therefore results should be treated with caution.

#### *English as an additional language (EAL) sub-group analysis*

We considered whether there is evidence of differential treatment effects between EAL and non-EAL pupils. Similar to the FSM sub-group analysis, we added to the primary and secondary analysis models an interaction term between the treatment variable and membership of the EAL sub-group. For all outcomes, the estimates for this interaction term are small and not statistically significant, providing little evidence of a differential effect between EAL and non-EAL pupils. In addition, we used the primary and secondary analysis models to estimate separate treatment effects for the EAL and non-EAL sub-groups. For both sub-samples and all outcomes, the treatment effects are small and not statistically significant. Table 18 summarises these results.

Table 18: English as an additional language (EAL) sub-group analyses

Outcome	Sub-group	Interaction term Hedges' <i>g</i> (95% CI)	Interaction term <i>p</i> - value	Sub-group total <i>n</i> (intervention, control)	Sub-group treatment Hedges' <i>g</i> (95% CI)	Sub-group treatment <i>p</i> - value
PIRA	EAL	0.07	0.62	471 (230, 241)	-0.02	0.83

		(-0.2, 0.34)			(-0.19, 0.16)	
	non-EAL			349 (180, 169)	-0.09 (-0.3, 0.12)	0.41
RBS	EAL	-0.02 (-0.29, 0.24)	0.87	456 (226, 230)	-0.05 (-0.23, 0.12)	0.55
	non-EAL			354 (178, 176)	0 (-0.21, 0.2)	0.96
WSEM3	EAL	0.1 (-0.19, 0.39)	0.5	443 (210, 233)	0.12 (-0.07, 0.3)	0.22
	non-EAL			326 (166, 160)	-0.02 (-0.25, 0.21)	0.85
SSIS	EAL	-0.03 (-0.34, 0.28)	0.84	451 (230, 221)	0.02 (-0.18, 0.21)	0.85
	non-EAL			294 (152, 142)	0.05 (-0.19, 0.29)	0.69

*Randomisation inference robustness check*

We also carried out alternative statistical inference for the primary and secondary analyses using randomisation inference to provide useful information on the extent to which there is variation between these different approaches to statistical inference. As seen in Table 19, the obtained results were not substantially different from the overall impact evaluation results.

Table 19: Sampling and randomisation inference p-values

Outcome	Effect size	Sampling inference $p$ -value	Randomisation inference $p$ -value
PIRA	-0.05	0.93	0.48
RBS	-0.04	> 0.99	0.51
WSEM3	0.05	0.47	0.42
SSIS	0.03	0.71	0.69

Note.  $p$ -values for the two primary outcomes (PIRA and RBS) are adjusted for multiple comparisons using the Bonferroni correction.

#### Clustered standard error robustness check

Although this is a pupil-randomised trial, pupils were nevertheless naturally clustered within schools, as confirmed by the observation of a non-zero ICC for each outcome (see *Estimation of ICC* below). Such clustering can inflate standard error estimates and increase the risk of false positives in statistical testing. This does not pose a risk to the current findings because the point estimates are already close to zero and these are not affected by clustering, but for completeness we report in Table 20 results from the primary and secondary analyses obtained using standard errors clustered at the school level. This table shows that the obtained results were not substantially different from the overall impact evaluation results.

Table 20: Additional analysis of primary and secondary outcomes using standard errors clustered at the school level

Outcome	Total $n$ (intervention, control)	Hedges' $g$ (97.5% or 95% CI)	$p$ -value
PIRA	821 (411, 410)	-0.05 (-0.2, 0.1)	0.88
RBS	811 (405, 406)	-0.04 (-0.26, 0.18)	>0.99
WSEM3	770 (377, 393)	0.05 (-0.1, 0.2)	0.46
SSIS	746 (383, 363)	0.03 (-0.22, 0.28)	0.88

Note.  $p$ -values for the two primary outcomes (PIRA and RBS) are adjusted for multiple comparisons using the Bonferroni correction and 97.5% CI are reported for the treatment effect size.

#### Estimation of effect sizes

As previously outlined, effect sizes for all analyses described above are calculated using Hedges'  $g$ . These have been presented in Tables 14 and 15 for the primary and secondary outcomes. Table C1 in Appendix C contains additional information used in the estimation of effect sizes (e.g., the standard deviations). None of the effect sizes estimated for the primary and secondary outcomes are of a substantive magnitude, nor are they statistically significant.

**Estimation of intra-cluster correlation (ICC)**

As this is a pupil-randomised trial, the ICC is not factored into the main analysis model. Nevertheless, as it may be relevant for future research, we report the ICC at the school level for each of the primary and secondary outcomes (see Table 21).

Table 21: ICC estimates for each outcome

Outcome	PIRA	RBS	WSEM3	SSIS
Intra-cluster correlation (ICC) estimate	0.13	0.08	0.03	0.02

## Implementation and process evaluation (IPE) results

This section describes the findings of the IPE, which are based on data collected from case studies and the supplementary post-intervention survey for all participating schools (see *Methods* for more details). The subsections are structured by theme and not research questions, in line with best practice qualitative reporting (Nowell et al., 2017). The research questions addressed by each subsection are outlined in Table 22. The first subsection provides contextual information pertaining to the case study schools' motivation to implement SB, as well as information on usual practice in relation to the support provided to pupils with speech and language difficulties. The second subsection outlines the factors that influenced the implementation and delivery of SB, including those influencing fidelity; that is, whether it was implemented as intended by the program developers. The final subsection describes the perceived outcomes and mechanisms for how SB brought about change.

Table 22: Overview of the research questions addressed in each subsection

Subsection in <i>IPE results</i>	Research questions
Context	RQ8 (Programme differentiation). What other support do the pupils access to support their communication in both control and treatment?
	RQ3 (Quality). How was the quality of the intervention perceived by teachers, senior leaders and TAs?
Implementation	RQ1 (Fidelity). In what ways was the programme implemented? What are the barriers and facilitators of delivery? In particular: <ul style="list-style-type: none"> <li><b>a</b> SLT 'buy-in';</li> <li><b>b</b> delivery of training and resources – whether it appears to be effective in ensuring that teaching staff understand the aims and main features of the intervention;</li> <li><b>c</b> delivery of the intervention – whether it appears to facilitate children's engagement.</li> </ul>
	RQ2 (Responsiveness). To what extent did the schools engage with the intervention in line with the intervention aims?
	RQ4 (Implementer support system). How was the knowledge of the arts practitioners delivering the intervention integrated with the knowledge of teaching staff involved? (Implementer support system)
	RQ7 (Adaptation). To what extent is the programme adapted by DPs and schools?
	RQ11 (Implementation environment). To what extent do school facilities affect the intervention?
Perceived mechanisms and impact	RQ5 (Mechanisms). What are the mechanisms that are taking place in the intervention and to what extent are they bringing about change?
	RQ9 (Reach). To what extent does the intervention affect the targeted children's classroom engagement and learning, particularly around engagement and communication?
	RQ10 (Implementer characteristics and context). To what extent does the intervention affect the TA, their role in school and with the teacher?

## Key IPE findings

- *Context and usual practice:* All trial schools surveyed indicated they had other supports available for pupils with speech and language challenges. Ninety percent of survey respondents indicated that in-class adult support was routinely available.
- *Adherence and compliance:* Most schools reported high adherence to the intervention, as indicated by the fact that 86% of survey respondents reported completing 20 or more SB sessions during the school year. Also, of pupils for whom we had session attendance data, 88% received at least 16 of 24 sessions (the minimum compliance threshold).
- *Fidelity:* Data gathered through interviews, a survey of all trial schools and observations provided evidence that fidelity to the general structure of the sessions (as outlined in the *Intervention* section of this report) was maintained with minimal modifications.
- *Implementation:* There were three key factors that influenced the implementation of SB with fidelity: (i) SLT support with securing the required resources; (ii) the SB facilitators acting as role models and partners in delivery; and (iii) the characteristics of the pupils in the SB sessions.
- *Mechanisms:* Based on data gathered, the logic model was updated with the removal of two mechanisms and the addition of one, which resulted in a final set of (i) creating a safe, non-judgemental and inviting space, (ii) use of 'I wonder' mode, and (iii) use of repetition and routine (see *Perceived mechanisms* in this section, and *Outcomes* in Figure 6).

*Perceived outcomes:* Members of the SLT, TAs and DPs in the case study schools generally reported feeling positive about the intervention and reported noticing positive changes in pupils' oral communication skills, self-efficacy and self-regulation. However, the perceived impacts on pupils' literacy skills were limited.

## Context

Staff in the case study schools reported being motivated to adopt SB with the aim of accelerating the learning progress and attainment of pupils with speech and language difficulties. This aligned with the key challenge reported by staff – of finding the best ways to support pupils with 'low starting points' (SLT member 05) and different needs, including speech and language difficulties. In some schools, staff reported that the reason for pupils having less developed speech and language was related to pupils having special educational needs (SEN) or because English was an additional language (EAL). At other schools, staff associated having greater numbers of pupils with lower starting points with the level of deprivation locally and the limited support they received for learning at home.

*'We are very diverse... I think there's about 40 languages spoken, different ethnicities. We've got a lot of children coming and going, so your class is changing all the time. We have got a lot of children who are new to English, quite a lot of SEN.'* (SLT member 03)

Members of school SLTs felt that SB would be valuable because it was a bespoke intervention tailored to the needs of pupils who were perceived to have speech and language difficulties. Staff liked SB because they felt that it was well-researched and used a more creative and kinaesthetic approach to learning, in comparison to typical pen and paper 'sit-down interventions' (SLT member 05). Furthermore, the size of pupil groups participating in SB were much smaller than a typical class size, which they felt would be beneficial to pupils as it would give them an opportunity to contribute more often and would give them more time to develop at their own pace. Positive views of SB were reinforced through hearing positive feedback from other schools about pupils' progress in SB and the SB taster session. The taster session was seen as particularly valuable because it allowed the school's SLT and other staff to better understand what SB involves and to envision how and which types of pupils might benefit.

*'One of the challenges that we face locally is that language levels and communication levels are very low on entry for the vast majority of pupils, and feedback from other schools in the area piloted the project was positive, so we thought we would give it a go.'* (SLT member 04)

*'I played some of the activities, so I was in role and things like that, so you had a first-hand experience of what the sessions would look like.'* (SLT member 01)

## Usual practice

School staff in the six case studies described the existing support available to pupils with speech and language difficulties. This included adapting class teaching and course materials, all teachers and TAs within the school being trained in the Elklan approach, and specific pupils being offered one-to-one or small group support with learning mentors or a speech and learning therapist. This was reflected in the survey data, which showed that all respondent schools had some type of support available, but the type of support varied (see Table 23).

Table 23: Participant responses to survey question #7.1: 'Outside the Speech Bubbles programme, what kind of support is offered to pupils with speech, language and communication challenges in your school?'

Range of support offered to pupils with speech, language and communication challenges	
Response	Number (%) answered 'Yes'
In-class adult support	19 (90%)
Out-of-class adult support	15 (71%)
Adaptation of class teaching and materials	11 (52%)
Other (please specify):	
• Speech and language therapist	3 (14%)
• Various intervention groups	2 (10%)
• Staff training from educational psychologist	1 (5%)
• Specialised SEN teacher	1 (5%)
• Support from external agencies	1 (5%)
No support is offered	0 (0%)

N.B. Total percentage is greater than 100% because respondents could select more than one answer.  $n = 21$ .

The most common support provided, as reported via survey by staff in 90% of the schools, was the in-class adult support that was offered to pupils with speech and language difficulties.

*'No matter what their learning style is or what their particular need is or where they've come from or what kind of home life they have, we work really, really hard to make sure everyone can access their learning.'* (SLT member 06)

Data is not available pertaining to whether there were any changes to the existing provision of support in the schools during the year the intervention was delivered.

## Programme implementation

This evaluation examined two implementation dimensions – compliance and implementation fidelity; the former was assessed using quantitative data and the latter was examined using data from the case studies (see *Methods* for more details). Specifically, compliance was conceptualised as a binary indicator of whether the pupil received a sufficient amount of the intervention to be considered as having received the intervention (i.e., treatment). For SB, pupil-level compliance in this trial was defined as having attended at least 16 of the 24 SB sessions (66%) prior to the collection of outcome data. In comparison, implementation fidelity was conceptualised as how the way in which the intervention was implemented in practice compares to the intended implementation of the intervention as described in the

*Intervention* section in this report. Case study data was used to examine the variation in how the intervention was implemented and any adaptations made in the case study schools, alongside identifying barriers and facilitators to implementing the intervention with fidelity.

## Compliance

Most schools demonstrated high adherence to the intervention, as indicated by the fact that 86% of survey respondents reported completing 20 or more SB sessions during the school year. All schools demonstrated sufficient adherence, which is defined as having delivered 16 or more sessions. The three staff who indicated delivering fewer than 20 SB sessions reported finding it challenging to prioritise SB relative to other learning objectives and whole-school training that conflicted with the delivery of the SB session. Further, in terms of pupil-level compliance, 87.2% of pupils (from 23 out of 26 schools from which we received the data) attended 16 or more sessions. This echoes a reflection from one of the DPs that sometimes attendance was a problem for certain pupils, such as for one pupil who missed too many sessions and did not return to SB. Our inference is that this issue might relate to the general challenges experienced by some schools – particularly those that have many pupils who are ‘coming and going’ (Teacher 03) and living in areas of high deprivation.

## Fidelity

Based on the interview, survey and observation data, there is evidence that fidelity to the general structure of the sessions (as outlined in the *Intervention*) was maintained with minimal modifications. All interviewed DPs reported making only minor modifications to the structure and approach, with the aim of tailoring it to the needs and interests of the pupils in the session. Some staff said that they brought in technology, props and created visual cues (e.g., a visual timetable of activities on the wall) in order to ensure all pupils were able to participate. For instance, one of the TAs brought in an iPad to show pictures, with the aim of supporting a pupil who is learning English with comprehension of the sessions’ content. Other staff reported gradually introducing new elements to existing activities, or new activities, to ensure that pupils continue to develop their skills and remain engaged.

*‘[SB] is quite standard, but you have to work with who is in the space... maybe one group might have two or three autistic kids. That needs something doing in a very different way or something more sensory, then obviously I adapt and make changes there.’ (DP 02)*

## Factors influencing implementation

There were three key factors that influenced the implementation of SB with fidelity:

- 1 SLT support with securing the required resources;
- 2 the SB facilitators acting as role models and partners in delivery;
- 3 the characteristics of the pupils in the SB sessions.

### **SB delivery needs SLT support with securing the required resources**

SLT support was important for ensuring that appropriate resources were identified and made available for SB sessions throughout the year. The key resources required included having a room in which the sessions could take place each week, and ensuring that there was a consistent TA available to co-facilitate the sessions. One of the elements of fidelity to the SB model is selecting an appropriate space, which includes having sufficient space for pupils to move around, with quiet surroundings and minimal decorations to minimise distraction (as detailed in the *Intervention*). In the six sessions that were observed, the spaces used were appropriate in terms of being quiet, relatively bright and open, with minimal distractions. This was echoed by DPs and TAs in interviews, as well as in the survey responses, which showed that 95% of respondents felt they had the space required for SB. However, some schools had difficulty with finding and securing a suitable space throughout the school year, due to limited space in the school in general and disruption caused by renovations. As a result, SB sessions moved to different rooms one or more times in some schools, which was viewed as having the potential to undermine the sense of familiarity for pupils.

Consistency in delivery also affected some schools due to the fact that they found it difficult to find a TA to participate in SB every week (for which the TA needed to leave their class), and challenges in finding a suitable time for the

sessions. For instance, based on survey and interview data, there were a few occasions when the sessions needed to be rescheduled due to other activities taking place, such as whole school training and school trips.

*'We had some building work going on with the roof, so this was used as a classroom and we had to go wherever we could, and we were moved around. That really disrupted [the sessions].'* (SLT member 05)

Where there were issues with resource availability, a supportive SLT played a significant role in helping to resolve them. The SLT in case study schools was described by TAs and DPs as supportive, helpful and responsive to queries. In schools that found it challenging to obtain a room, the SLT resolved this issue by communicating to other school staff the importance of having a consistent space and being firm about this need. This echoes the survey findings across all participating trial schools in which none of the survey respondents indicated a lack of SLT support as being an issue.

*'In this school I felt very supported by the staff, and with the challenge with the room that was solved straightaway.'* (DP 06)

### **SB facilitators as role models and partners in delivery**

The data from observations and interviews with DPs and TAs indicates that the SB facilitators (DPs and TAs) influenced how SB was delivered in terms of the fidelity to the SB ethos and values, which place an emphasis on creating a safe and nurturing environment. This was enabled by the SB facilitators being:

- 1 partners with complementary roles in the delivery of the sessions;
- 2 being role models for the pupils.

#### *(i) SB facilitators as partners in delivery*

Both TAs and DPs felt that they had a good collaborative working relationship in which they supported each other. Their roles were complementary: the DP's primary role was to lead the session, while the TA focused on managing the pupils' behaviour, as well as supporting and encouraging participation. The roles that the SB facilitators took on aligned with their knowledge and expertise. DPs were the drama experts, with all DPs reporting being in the field for a long time (e.g., over 25 years as a DP). They also had experience with delivering SB for between two and six years, following an extensive training programme. Overall, 95% of the survey respondents reported being either very satisfied or somewhat satisfied with the DP's delivery of the SB sessions. The DP's expertise combined with their role in leading the SB sessions was central to maintaining fidelity to the structure of the sessions and activities used (as described above), with minimal adaptations made. This compares to the staff co-facilitators, most of whom did not have any prior knowledge of SB or previous drama experience. The exceptions to this were one staff member who had previously co-facilitated SB for one academic year and another who had a professional drama background.

The TAs had an important role in managing behaviour, being an active participant in the SB sessions, and supporting pupils' participation, which was enabled by their knowledge of the pupils. In fact, some SLT members said that they chose those TAs to be the co-facilitators due to their familiarity with the pupils. This was valuable because the TAs were able to share with the DPs relevant details about the pupils needs, characteristics and personal circumstances, to help contextualise their behaviour. When pupils displayed recurrent disruptive behaviour, the TA and DP worked together to identify the most appropriate strategy for managing the behaviour and considered whether other adjustments could be made in the sessions.

*'After every session we'll sit there and he'll [DP] say little things like, this child was doing this, then I'll say "Well, that child is autistic." So it's worked really well.'* (TA 01)

*'We're a partnership. In terms of discipline, I will always say to my teaching assistant is that you are in charge, whether you decide there is an issue... she knows that child outside of the space and bringing in that context.'* (DP 03)

On a few occasions in which the TAs did not know the pupil well, the SLT was brought in to help resolve the issue.

Further, the TAs role evolved over time as they became more confident with drama and familiar with SB. Observations of SB sessions towards the end of the year showed evidence that all TAs in case study schools confidently

participated in the sessions. Specifically, some TAs began to lead activities and provided more support with running the sessions. This was echoed in the interviews with the TAs who had no drama experience and initially reported being a bit nervous about helping to deliver SB sessions, but said they gradually developed their confidence following the induction session. In the survey, all TAs reported feeling like they understood the purpose of SB, could co-deliver basic SB activities and could keep the pupils engaged (as indicated by 100% of responding TAs selecting 'strongly agree' or 'agree' in the survey).

#### *SB facilitators as role models in delivery*

A key role of the SB facilitators was to be role models for the pupils in terms of the programme's ethos and values. This manifested in how the pupils were encouraged to participate and the approach to behaviour management, as indicated by the interview and observations data.

Initially, TAs' active participation in SB activities was central to encouraging pupils to participate. This is because the pupils were still adjusting to the new adult (i.e., the DP), as well as the new environment, activities and expectations. Thus, pupils looked to the TA for comfort and guidance, especially as it related to being willing to try new things and make mistakes. This was integral to the process of pupils exploring how they would like to express themselves, especially those who were viewed as quiet. TAs were able to empathise with the pupils, because drama and SB was a new activity for them too.

*'[The TA's role] is very much being a member of the group in a way, and also modelling stuff... [it] is really important that the children see that this adult is happy to have a go at things and try things out and happy to make mistakes.'* (DP 04)

*'I kind of explained to them that this is all new to [me] as well, so we have learnt and grown together.'* (TA 02)

Based on the interview and observation data, there is evidence that all pupils received encouragement and positive feedback about their contributions and participation. However, TAs and DPs said that more reserved pupils, in particular, needed more support and encouragement at the beginning of the year. This included verbal encouragement, as well as having a familiar adult do an activity with them when it was their turn. Over time, more reserved pupils were able to build their confidence to participate independently.

*'When she wasn't happy to step into the square you know, the member of staff came with her and it built her confidence up and you could see that, that the next time she was much more confident in doing that and participating independently and it was very supportive.'* (Teacher 02)

Generally, the more boisterous pupils and those who had behavioural issues needed reminders and encouragement to behave appropriately. In line with the SB ethos, the TAs did so in a calm and gentle manner, as well as providing praise when the pupils' behaviour changed. Specifically, the facilitators used the SB rules as their key reference. This was helpful because the pupils were familiar with the SB rules as they were recited at the beginning of every session. In both the interviews and observations, there was evidence of TAs using the rules as a reference when managing behaviour, as well as their manner for doing so.

*'This involves encouragement to take part in a particular way (e.g., "in Speech Bubbles we take turns / do good listening don't we?"), rather than telling them to stop doing something. The manner is kind but firm. The interjections from the TA are quiet and unobtrusive.'* (Observation notes school 05)

#### **Impact of characteristics of pupils and group dynamics on delivery**

The characteristics of pupils in the SB sessions influenced the delivery of the programme. The importance of selecting the right pupils was emphasised by schools, particularly those that experienced challenges due to having pupils that they felt were less suitable for SB. In these schools, the SB facilitators reported having difficulty with managing some of the pupils' behaviour, which was sometimes disruptive to the sessions and required one of the facilitators to spend more time focusing on that particular pupil. The views about which pupils might be more or less suitable varied in terms of age, presence of behavioural issues and general disposition. However, some DPs and TAs in the case study schools felt that it was important to consider not only the characteristics of the pupil, but also how this might affect the group dynamics.

*'It can be challenging because there's all sorts of children's behaviour, then you will have children who don't say anything or perhaps are selectively mute. Then you have got children who are really loud and interrupt, and the children are up and down, up and down.'* (TA 01)

In terms of age, staff in some schools felt that the older Year 2 pupils benefited most from the SB sessions, while others felt that these pupils were less suitable because of what was described as 'Year 2 syndrome' (TA 03). That is, some pupils in Year 2 were described as feeling too old for some of the activities, because they were perceived to view them as silly. As a result, their level of engagement could fluctuate. In one case, the TA and DP noticed one pupil, who would occasionally disengage from some activities, acted as a peer role model for negative behaviour for a few other pupils.

*'She's got like Year 2 syndrome, where it's like she's almost too old for it and she knows that this is just pretend... So if she is kind of having an off day, they might look to her and think, I'm going to do that as well, and copy her.'* (TA 03)

Some interviewed DPs and TAs reported that it was challenging to manage the behaviour of disruptive pupils and difficult to mitigate the impact of this on the rest of the group. As a result, some felt that such pupils were less suitable for SB, while others felt that these pupils could still benefit. However, some interviewed DPs and TAs felt it was not just about the individual pupils' characteristics but rather how their characteristics impacted the group dynamic. One DP, for example, recommended that pupils with different dispositions (e.g., reserved and loud / disruptive) should not be separated because this would create an artificial environment. This would not be beneficial in the long run, as pupils might continue having difficulty with speaking and participating in a classroom setting in which there are pupils with different dispositions. Instead, they suggested that deciding whether a pupil is suitable should be done on a case-by-case basis, taking into account the potential level of disruption they might bring. This issue reinforced the importance of having a TA co-facilitator who knows the pupils and is skilled in managing behavioural issues.

*'Early days was he dominating so much that he took away from some of these sessions previously and was he an appropriate candidate. So, I think if we were doing it again we would probably look at a little bit of that.'* (SLT member 02)

*'I think it can be destructive, depending on the level of disruption, and we have to make a joint decision with the TA in what that gets to and whether that becomes an issue. But I also think they have to manage that in class. So to completely create an artificial environment when you only put disruptive children together and only put quiet children together would not necessarily benefit them in the long run.'* (DP 03)

## Perceived mechanisms and outcomes

This section describes the mediating mechanisms and perceived outcomes of SB, as well as the changes to the original logic model that have been made in accordance with the evidence from the interviews, observations and survey of all the trial schools. The first subsection focuses on the pupil mechanisms and outcomes, while the second subsection focuses on the TA mechanisms and outcomes. At the end of this section, a revised logic model is presented (Figure 6).

### Pupil mechanisms and outcomes

In the original logic model, four key mediating mechanisms were specified:

- 1 creating a safe, non-judgemental and inviting space;
- 2 use of 'I wonder' mode (further explained below);
- 3 positive relationship between the TA and DP;
- 4 an encouraging teacher.

These mechanisms are expected to lead to the outcomes of increased self-efficacy and self-regulation, and, over time, to positively influence the longer term outcomes of oral and written communication skills. The first two mechanisms were supported by the evidence from the IPE. However, the IPE data did not contain evidence in support of the latter two mechanisms, so they have been removed from the logic model. The third mechanism – the relationship between

the TA and DP – was found to be a factor influencing SB delivery (as detailed in the *Implementation* and *Fidelity* sections), rather than a mechanism affecting pupil outcomes. There was also no evidence in the IPE data supporting the fourth mechanism – the role of ‘an encouraging teacher’, which may have been due to teachers reporting that they had limited to no engagement with SB. Furthermore, the 13 additional mechanisms that were included in the original logic model (Figure 1) have been removed in the revised logic model (Figure 6). The analysis of the IPE data provided evidence that some of these mechanisms – namely, participation, engagement, fun, trust, resilience, peer relations, turn taking, context-appropriate behaviour, improved attention and creativity / imagination – were sub-components of the key mediating mechanisms presented in the revised logic model. There was no evidence in the IPE pertaining to the outstanding additional mechanisms that were included in the original logic model, specifically refined motor skills, narrative structure and purpose writing.

In the revised logic model, repetition and routine were added as an additional mechanism, alongside the existing mechanisms of creating a safe, non-judgemental and inviting space, and use of ‘I wonder’ mode (which is described below). The IPE did not find evidence of increased pupil self-esteem, and therefore this has been removed as an outcome from the logic model. The findings for each of the mechanisms and outcomes are presented below.

*Mechanism: Creating safe, non-judgemental and inviting space*

The first mechanism involves the creation of a safe, non-judgemental and inviting space, which is expected to improve the pupils’ day-to-day experiences of learning to communicate. The interview and observations data indicate that this space was created through having the SB facilitators (TAs and DPs) continuously reinforcing that there were no right and wrong answers or responses. For instance, the facilitators did not change the pupils’ stories, even when some elements of the narrative were viewed by the co-facilitators as being harder to follow. This helped pupils to feel safe to try new ways of expressing themselves, because they knew that it was okay to make mistakes and ask questions if they did not understand. This was contrasted by staff with a pedagogical approach that prioritises and encourages pupils to get the *right* answers.

*‘But they will cover that [not understanding] up in the class context to just get along or they might just get told off, or they fear being told off. I think it allows them that space to start maybe asking questions and say, I don’t know that word and I don’t understand.’ (DP 03)*

The interviewed TAs and DPs reported that the SB rule of being gentle with each other also reinforced the non-judgemental ethos. This allowed pupils to experiment with expressing themselves without fear of being judged by the other pupils in the session. As a result, the SB co-facilitators reported that the pupils became braver and more willing to take ‘*risks in their learning*’ (SLT member 06).

*‘Early on we established that we don’t comment on what other people do... if someone makes a mistake, we really want there not to be a culture where children are scared to make mistakes, because they have to take risks in Speech Bubbles for them to make that progress.’ (DP 01)*

As described by the interviewed TAs and DPs, the non-judgemental ethos was also reinforced by providing an opportunity for all pupils to participate and valuing all participation equally regardless of how much they contributed. While all received the opportunity to participate, they did not have to participate if they did not want to, and their choice was respected. This is echoed by the data from interviews and observation sessions in which some pupils were louder and others quieter when they spoke. In terms of enabling participation, this included prompting them to contribute, providing verbal encouragement and additional time to think before contributing. This allowed pupils of all attainment levels and temperaments to participate in a way that works best for them and at their own pace.

*‘The quieter children are encouraged to take part and given plenty of praise. Their approach is quieter and more hesitant but they are not told that they have to do it more loudly or confidently.’ (Observation notes school 06)*

*Mechanism: ‘I wonder’ mode*

The second mechanism is the ‘I wonder’ mode, which is a form of indirect questioning used by the practitioner to elicit pupils’ ideas. The ‘I wonder mode’ was fostered by providing pupils with the opportunity to express themselves creatively. The interviewed TAs and DPs described how SB allowed pupils to use their imagination to develop content for their stories, as well as use their body to act out the stories. As a result, they were able to take on different

characters, experiment with using different voices and body movements to express themselves. Staff highlighted that this approach was particularly valuable because it accommodated pupils who had different types of learning needs.

*'[SB] really gives them a chance to express themselves and act out fun things and just explore the different ideas with no right or wrong answers' (TA 03)*

*'[The SB session had a] story about a princess and a flower. DP encourages students to add new dimensions to performances (e.g., everyone adding sound effects) ... or to speak whilst acting (e.g., what might the princess say at this point etc.)' (Observation notes school 02)*

In SB, learning happens while having fun through games, which was described by staff as being different to how learning typically happens in school. As a result, pupils would return to class from SB looking refreshed, given that they had the opportunity to move around and release their energy, as noted by some staff. This was also echoed in the discussions with the pupils, many of whom loved SB because they found it fun and different from the work they do in class. Overall, most staff thought that SB was somewhat or very engaging for pupils, as reported by 95% of the survey respondents.

*'One of the boys said he liked the "I can do this" game because he can spin around on one leg, and this is something he can't do in the normal classroom.' (School 05 notes from discussion with pupils)*

*'[Pupil] liked SB more than her normal lessons because in lessons they "do lots of work" and in Speech Bubbles they have "lots of fun".' (School 02 notes from discussion with pupils)*

#### *Mechanism: Routine and repetition*

The third mechanism relates to one of the key attributes of the SB sessions, which is the element of routine in terms of how the sessions are structured and the repetition of the activities. The SB co-facilitators said that it was important for the sessions to happen at the same time each week, in the same space and with the same facilitators. This was described as acting like a *'comfort blanket'* (TA 06), which helped pupils feel safe and secure. Similarly, each session had the same structure in terms of the order and types of activities, such as the Good Morning / Hello game, and chanting the SB values / rules at the beginning of the session. A routine ensured that pupils knew what would happen next, while repetition of activities enabled them to explore how they would like to express themselves with the knowledge that they could try again in the next session. Over time, pupils' self-efficacy was described as increasing as they became more familiar, comfortable and confident in engaging in those activities and developed a sense of competence. This highlighted the importance of fidelity to the structure of the SB sessions, which was reflected in the DP accounts of only making minor, gradual changes to activities over time, if the pupils were ready for a challenge.

*'It has been very consistent with going out and that's what the children like, the consistency of something happening. They know that when it's Thursday it's Speech Bubbles day, so they are quite excited about going.' (Teacher 05)*

Another element of repetition was having a consistent set of SB rules, which included taking turns and raising their hands. Pupils were reminded of these rules at the beginning of the sessions and referenced by the co-facilitators as part of their approach to behaviour management. The turn-taking element was described by the SB co-facilitators as important because it led to pupils' having trust that they would get their turn to have their voice heard during every session, which was also made possible by having a small group of students. This provided the quieter pupils a chance to speak, who might not otherwise get a chance to do so in the classroom, and calmed the more boisterous pupils.

*'So kind of you know children that are maybe a little bit in their shell and a little bit quiet. Because it was in smaller groups and because everyone would have the chance to be able to have their story told if they wanted to.' (Teacher 01)*

*'[Pupils] really do trust that whole circle method that you will get your turn, it's a really powerful thing that because I think it can calm people down knowing that they can have their go.' (DP 05)*

*Outcome: Self-efficacy and self-confidence*

In the case study schools, staff noticed increases in pupils' self-efficacy in terms of their contributions and participation in the SB sessions.<sup>62</sup> In the survey, 91% of the TA survey respondents reported seeing a 'somewhat positive' or 'very positive' impact on pupil's confidence. The extent of the impact ranged from 'speaking up more often' to taking a lead in some of the activities in the SB sessions. For some pupils, the impact of self-efficacy was seen in their increased willingness to contribute, and volunteering to speak up, as opposed to being prompted to do so. For others, such as the pupils who were selectively mute, starting to answer questions was reported by staff as a big improvement.

*'There was one particular boy that wouldn't contribute at all to my lessons at the beginning of the year. Whereas now, he puts his hand up. We even saw him put his hand up in assembly the other day and spoke in front of the whole school, which he would never have been able to do that at the beginning of the year.'* (Teacher 02)

*'One little girl used to sit there with her fingers up to her mouth and not speaking. Now she does, she joins in, she volunteers. So, it has helped. The children are developing and getting confident and they usually run the session.'* (TA 05)

The impact on self-efficacy was also noticed outside of the SB sessions in terms of participation in the classroom, playground and other school activities. For instance, the TAs and teachers reported that they were noticing more that the pupils who participated in SB were raising their hands and contributing to class discussions. In some cases, staff reported receiving feedback from parents, who noticed positive changes such as their child being more confident. Also, there was evidence that increases in pupil's self-efficacy positively impacted how they interacted with adults and their peers in the playground and in classrooms. This was echoed in the survey, in which 95% of respondents reported that they believed SB had a 'somewhat positive' or 'very positive' impact on pupil's social skills.

*'I know that I spoke to a parent who was glad that her child was doing Speech Bubbles because she has noticed a difference in her confidence.'* (TA 03)

*'You can definitely clearly see it from a social point of view on the playground... I've noticed more than once, those key children playing nicer and being more confident to make friends and just seemingly happier on the playground.'* (SLT member 06)

However, for some pupils, staff did not see changes in pupils' confidence, which was attributed to the fact that those pupils were already confident. For one pupil, the TA thought that they might have become over-confident, which was viewed negatively as it was causing them to become more inattentive in the sessions. Rather than the broader concept of self-efficacy discussed here, the impact evaluation focused specifically on creative self-efficacy and did not find any statistically significant impact on this outcome.

*Outcome: Self-regulation*

School staff and DPs reported noticing improvements in pupils' self-regulation, as a result of their participation in SB. This included improved listening and attention, turn-taking and generally, more appropriate context-specific behaviour, which was attributed to the SB rules and approach. The impact on self-regulation was noticed in the SB sessions, as well as in the classroom for some pupils.

*'When they do paired work you can often see him where he's got his ideas but he's now learning to take the turn and share that with other people, which is really good to see.'* (TA 04)

*'But he does come into class and he is much more respectful of others' ideas and opinions, and in general working in groups.'* (Teacher 01)

The DPs and staff said that the impact on self-regulation was particularly noticeable for the more boisterous pupils, who had a tendency to interrupt, not listen and not pay attention. Staff felt that these pupils became calmer, more able

---

<sup>62</sup> It should be noted that, when exploring the impact on self-efficacy, some participants used the term interchangeably with self-confidence. As such, in all claims, we have used the term used by the participant.

to wait to take their turn and gave other pupils a chance to speak more often. The session structure was perceived to have helped the pupils to trust that they would have their turn to contribute, alongside the activities providing an outlet for some of their energy.

*'Some louder, more confident students occasionally interrupted sessions or [were] keen to have their voices heard, DP reminded them that it wasn't their turn yet but made sure to come back to them later on.'* (notes from observation School 03)

*'But there is a boy in Year 2, who was so loud and so controlling ... He is so much calmer and he talks about his Speech Bubble strategies. He talks about how he needs to think about things first before he does them.'* (SLT member 01)

However, the impact on self-regulation was not seen for all pupils, particularly those described as having behavioural issues. In some cases, staff reported noticing that some pupils' behavioural issues were magnified as a result of participating in SB, which they believed might be due to the challenge of adapting to a new environment with new rules. In addition, some school staff and DPs acknowledged that developing new skills takes time, which might in part account for why the impact on some pupils might be less noticeable. For this reason, DPs recommended that it would be beneficial for some pupils to do more than one year in SB, if possible. This highlights the importance of considering pupils' starting point in terms of their usual behaviour and temperament, and gauging progress accordingly. This is consistent with the impact evaluation results, which show no statistically significant impact on pupils' social skills.

#### *Long-term impact on written and oral communication skills*

Pupil's increased self-efficacy and self-regulation were hypothesised to positively influence oral and written communication skills in the long term. Some staff reported already seeing positive changes in pupils' oral communication skills, in terms of how much and how they communicated. In the surveys, 86% and 81% of staff thought that SB had a 'somewhat positive' or 'very positive' impact on pupil's communication, and speech and language, respectively.

*'He came into Year 1 and his speech and language was quite limited because I know that he gets on well with Speech Bubbles, and it's tempting for me to say that's kind of helped him with his speech and language.'* (Teacher 04)

However, the reported impact on written communication skills and literacy was perceived to be limited, with only 38% of survey respondents reporting a 'somewhat positive' or 'very positive' impact on reading. In the few instances where impact was noticed and reported by staff, this was related to seeing improvements in pupils' handwriting and structure of writing, as well as pupils using new words.

#### **TA perceived mechanisms and outcomes**

The TA mechanisms have been updated, and outcomes for TAs have been added to the logic model in Figure 6. While data are not available to further comment about how they connect to the pupil outcomes, we believe that it is nevertheless important to acknowledge the perceived impact on TAs. In terms of mechanisms, the evidence indicated that TAs' knowledge and self-efficacy for supporting pupils with speech and language difficulties increased, in accordance with the logic model. This was self-reported by TAs in the interviews and echoed in the survey responses, in which 86% of staff co-facilitators felt that SB had a 'very positive' or 'somewhat positive' impact on their understanding of how children learn. However, the evidence does not suggest that TAs felt that they had become speech and language specialists; thus, this has been removed from the logic model. Instead, by participating in SB, some TAs became more interested in learning about how to best support pupils with speech and language difficulties.

TAs gaining a new perspective on how children learn was added as a new mechanism based on data from the survey and case studies. The survey findings indicate that 86% of respondents felt that participating in SB had a 'very positive' to 'somewhat positive' impact on their understanding of how children learn. The new perspective included having a better understanding of the importance of routine and giving all pupils a chance to have their say. This new perspective led the TAs to apply the SB thinking and strategies in practice outside of the SB session, including in the classroom and drama activities they were involved in.

*'So, getting them to talk about what they're doing and what they're thinking and what they're going to say and taking note of that more than she did... I think she's giving everybody more chance to have their say now as well. So definitely a positive effect on that.'* (Teacher 05)

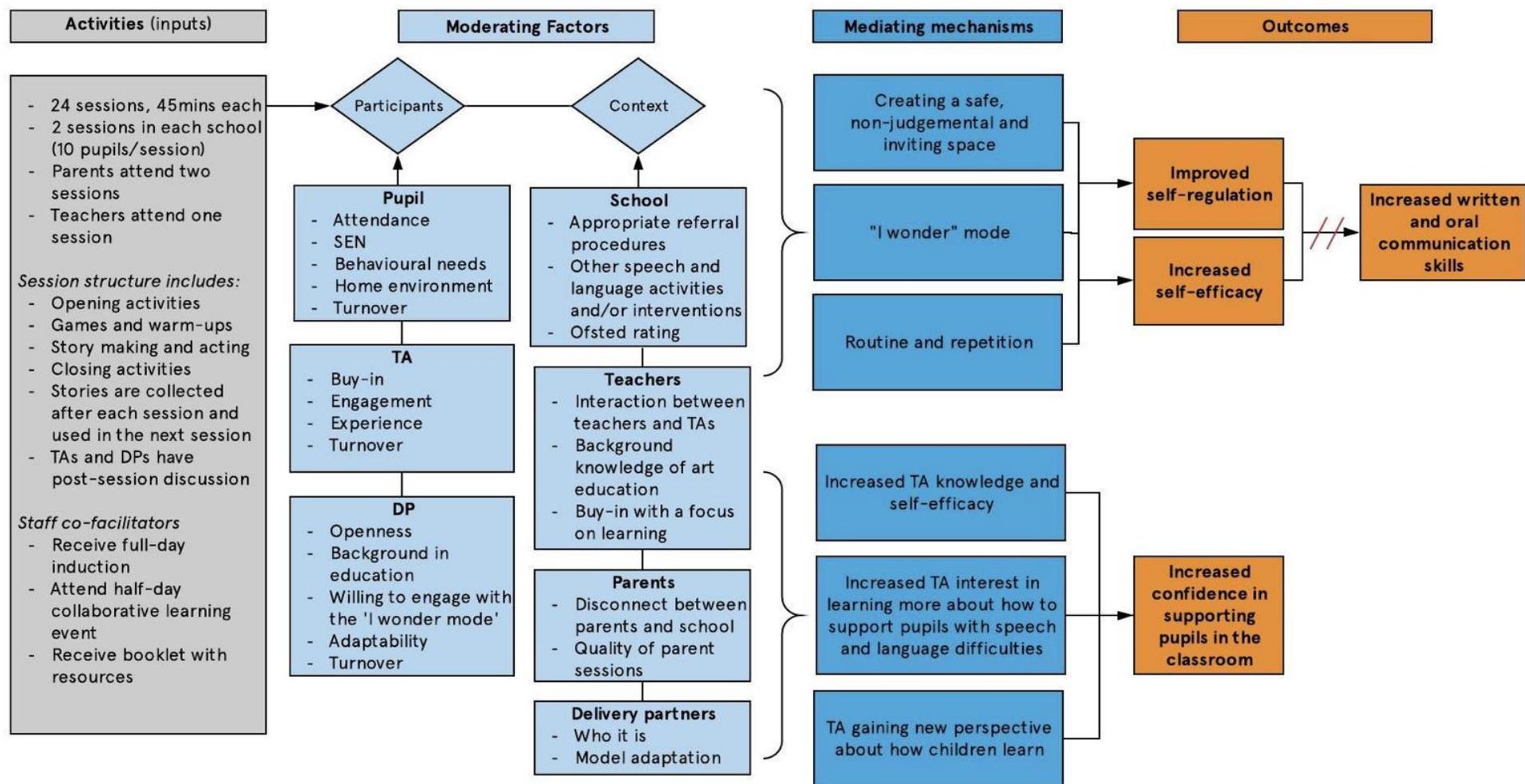
*'I also used it, it was during a lesson time where we were doing English, the Gingerbread Man song and I use that sort of structure again where it was okay, how would you move and then stop and go and things like that. So, taking little bits of that to use in my own you know drama things but also different areas of the curriculum.'* (TA 04)

Also, there is evidence that the TA mechanisms led to some TAs having increased confidence in the classroom in terms of supporting and interacting with pupils, which has been added to the logic model. This was captured in the survey, with 76% of the SB co-facilitators reporting having increased confidence in the classroom. In addition, it was also noticed by some teachers and SLT members, who reported that their TAs have made suggestions to the teacher about using strategies that they learnt in SB.

*'So the drama skills that she is using in Speech Bubbles she has taken it through to her classroom and suggesting to the teacher have you thought about this, this is what we do and it helps the children and understands the language a bit more.'* (SLT member 03)

The TAs' self-reported improvement in confidence and self-efficacy and changes in classroom practices do cause some concern about spill-over to control group pupils. This issue is addressed in the conclusion.

Figure 6: Revised SB logic mode



**Elements that were removed from the original logic model (due to lack of evidence)**

- **Mediating mechanisms:** Encouraging teacher; positive relationship between the TA and DP; TA's becoming speech and language specialists;
- **Outcome:** self-esteem

\*We have not removed the primary outcome of increased written and oral communication; however, we have indicated with red lines that we did not find evidence that these outcomes were impacted by the programme.

## Cost

Delivery of the Speech Bubbles intervention cost approximately £540 per school for the year it was delivered. The direct cost to schools for the programme is £3000, but this is often subsidised by additional funding sourced by London Bubble or the regional theatre companies delivering the programme. Programme fees for schools participating in this evaluation were £500 per school, as the fees were subsidised by the EEF. To calculate the total cost per pupil over three years, we assumed the number of pupils would be 20 pupils in each year, as per the programme model. We also assumed no subsidisation of programme fees in Years 2 and 3. Based on these assumptions, **the total cost per pupil per year over 3 years is £110** (see Table 25). The cost breakdown is set out in Table 22 (all figures are rounded to the nearest pound).

Table 24: Cumulative costs of Speech Bubbles (SB) (assuming delivery over three years)

Year	Cost (£)	Number of pupils (cumulative)	Cost per pupil	Cost rating
Year 1	£540	20	£27	£
Year 2	£3556	40	£89	££
Year 3	£6572	60	£110	££

Source: Interviews with teachers in cost case study schools.

Table 25: Detailed costs of delivering Speech Bubbles (SB)

Item	Type of cost	Average cost (first year) (minimum, maximum)	Total cost over 3 years	Total cost per pupil per year over 3 years
Materials and printing	Ongoing	£16 (£0, £60)	£48	£1
Expenses (travel and subsistence to attend CPD)	First year	£24 (£0, £40)	£72	<£1
Programme fees paid to developer	First year	£500	£6500	£108
<b>Total</b>		<b>£540</b>	<b>£6620</b>	<b>£110</b>

Source: Interviews with teachers in case study schools.

## Training

To support the delivery of the intervention, schools allocated either a TA or a teacher to attend three days of training. In general, schools used internal staff at no additional cost to cover the days TAs or teachers were absent; only one school reported that they used a supply teacher to cover the days. In terms of staff time, training was estimated at 8 hours per school. TAs are encouraged to attend the training each year of implementation.

### **Preparation**

There was some additional administrative time associated with supporting the intervention. Cost survey respondents reported spending on average 1 hour discussing which pupils to refer to the programme (ranging from 0 to 2.5 hours). Teachers also reported that in the first year they spent approximately 9 hours per school preparing for the programme (ranging from 0 to 20 hours over the course of the year.) Assuming the same level of preparation is required each year, staff time for preparation over three years would amount to approximately 30 hours.

### **Delivery**

The recommended amount of staff time required for the delivery of the intervention is 45 mins per week, on average schools delivered 40 hours (or 5 days) per school per year. The reported hours of delivery ranged from 44–60 hours per year.

## Conclusion

Table 26: Key conclusions

### Key conclusions

1. Pupils who received the Speech Bubbles intervention had, on average, lower reading attainment scores (equivalent to one month's less progress) as compared to children in the control group. Pupils who received Speech Bubbles did not show any substantial difference in oral communication skills as compared to pupils in the control group. These are our best estimates of impact which both have a high security rating. However, as with any study, there is uncertainty around the result: the possible impact of this programme on both reading attainment and oral communication ranges from three months less progress to positive effects of two additional months of progress.
2. The impact evaluation did not find evidence the Speech Bubbles intervention had any effect on creative self-efficacy or social skills. The intervention did not have a differential impact on reading attainment or oral communication skills by eligibility for free school meals.
3. Members of the SLT, TAs and drama practitioners in the case study schools generally reported feeling positive about the intervention and reported noticing positive changes in pupils' oral communication skills, self-efficacy, and self-regulation. However, the perceived impacts on pupils' literacy were limited.
4. Overall, fidelity to the delivery model and dosage were high, based on data gathered through interviews, surveys, observations, and delivery records.
5. The key factors identified that influenced implementation and delivery of the intervention were: (i) the need to have SLT support to secure required resources (e.g., the same room each week); (ii) SB facilitators acting as role models and partners in delivery; and (iii) the characteristics of the pupils and associated group dynamics during the SB sessions.

## Impact evaluation and IPE integration

### Evidence to support the logic model

Changes to the mechanisms and outcomes elements of the original logic model (Figure 1) have been made in accordance with the evidence from the impact evaluation, survey and case study findings. For the mechanisms for pupil outcomes, two mechanisms – positive relationship between the TA and DP and an encouraging teacher – were removed, given that the IPE data did not contain supporting evidence. Instead, the IPE data indicated the importance of repetition and routine as a factor that is expected to lead to the proximal outcomes; thus, this mechanism was added to the revised logic model (Figure 6). For the perceived outcomes for pupils, data from the case studies and survey for all trial schools indicate that there is evidence that the intervention increased self-efficacy and self-regulation for some pupils; however, there was no evidence indicating that there was an impact on pupils' self-esteem, and thus this has been removed from the logic model. The data from the case studies and trial schools indicates that there is limited evidence that SB positively influences the longer term outcomes of reading and oral communication skills. The impact evaluation did not find evidence showing that the intervention improved pupils' reading and oral communication skills, social skills, or creative self-efficacy; we have updated the logic model to indicate that we have not found evidence to support that these outcomes were impacted by the intervention.

Also, changes were made to the mechanisms and outcomes for the TAs based on the IPE evidence. The mechanism of TAs becoming speech and language 'specialists' was substituted with 'TAs becoming motivated to further learn about how to best support pupils with speech and language difficulties'. In addition, there was evidence in the IPE that TAs gained a new perspective on how children learn and therefore this was added as a mechanism. Furthermore, the evidence indicates that these mechanisms lead to TAs having increased confidence with supporting pupils in the classroom, which was added as a new outcome for TAs.

However, no changes were made to the *Activities* and *Moderating factors* sections. In terms of the *Activities*, the evidence from the surveys and case studies indicates that the activities listed in the logic model were implemented in the schools. In terms of the *Moderating factors*, these were not revised, given that the case study data did not provide evidence to be able to confidently understand which and to what extent moderating factors impacted mechanisms and outcomes. Overall session attendance was high, which limited the impact evaluation's ability to assess whether attendance is an important moderating factor. The analysis in the presence of non-compliance did not provide evidence to suggest that attendance is a moderator of the outcomes we analysed. Similarly, analyses of differential treatment effects between FSM and non-FSM pupils and between EAL and non-EAL pupils did not provide evidence that either of these factors are important moderators of the outcomes used in this evaluation. However, we also did not

find evidence that the programme had an impact on the intended outcomes, so we are limited in the conclusions we can draw about the importance of these moderators.

## Interpretation

SB is an intervention that aims to use creative drama sessions to improve the communication and social skills of pupils who struggle in those areas. Though SB is a relatively new approach within the field of pedagogy, a number of existing studies have concluded that it is a promising intervention. These include a 2013 evaluation that found that a large proportion of teachers felt the intervention led to improvements in learning, speaking and listening as well as benefits to emotional regulation and conduct. Similarly, a 2016 evaluation found that children who participated in SB showed improvements in spoken language, storytelling and social interaction. This work suggests that SB has the potential to support pupils.

These previous evaluations of SB sit within a wider literature regarding the benefits of cultural learning programmes. This literature suggests that these approaches can go beyond aiding communication and social skills and are associated with improvements in more academically focused areas such as mathematics and literacy (Newman et al., 2010; Catterall, 2009; Catterall et al., 2012). Given this context, the current evaluation set out to assess the extent to which SB could make a measurable difference to reading skills and narrative oral skills (the primary outcomes), as well as increases in creative self-efficacy and social skills (the secondary outcomes).

The outcome of the evaluation was that no statistically significant effects were found for either the primary or secondary outcomes. The measures of reading and narrative oral skills both indicated that the average outcome for the intervention group was marginally lower than for the control group (but not statistically significant) and for creative self-efficacy and social skills the average outcome for the intervention group was marginally higher than for the control group, but again, the difference is not statistically significant. These findings remained the same when the analysis took into account levels of compliance, and also do not show any difference when looking at pupils in receipt of FSM or those for whom English is an additional language (EAL). Therefore, the evaluation does not support the hypothesis that SB can improve the outcomes measured in this study.

Given the promising nature of previous evaluations and the wider literature indicating that creative interventions can support learning, this finding is somewhat surprising. Unlike in evaluations of some interventions delivered in a real-world context, the lack of impact does not seem to be related to problems with fidelity. In the SB case, the data indicated that compliance was very high, with almost 90% of treated pupils attending at least 16 sessions and treated pupils attending a median of 22 out of 24 sessions. A high level of compliance is an indication that teachers, TAs and schools felt positive about SB, and this is supported by data from the school survey, which found that 86% of staff co-facilitators felt that the intervention had had a positive impact on their understanding of how children learn.

The positive impression of the programme on TAs was echoed by the fact that a large majority of TAs who completed the survey felt SB improved pupils' confidence, social skills and communication. At an individual level, some teachers and TAs described in qualitative interviews that they noticed that some pupils' behaviour changed, not only in SB sessions, but also in the classroom and playground more generally. These behavioural changes included speaking up more often, taking the initiative to communicate and exhibiting improved social skills such as turn-taking and context-specific behaviours. These impressions seem to run counter to the findings of the impact evaluation, which did not find evidence of impact on social skills. But it is important to bear in mind that improvements in individual cases that are (selectively) noticed by staff will not be sufficient to substantially change average impacts.

More in line with the findings of the impact evaluation, which did not find evidence of impact on reading attainment and oral communication skills, a much lower proportion of TAs in the survey (38%) reported that they felt the intervention had a positive impact on literacy. This was also reflected in the interviews, where staff felt they had noticed initial changes in this area but that these were limited and would need more time to become apparent. In accordance, some DPs and TAs felt that some pupils would benefit from doing another year in SB to continue developing the skills. It is worth noting that the intervention's theory of change shows that improvements in literacy and oral skills are mediated through improvements in psycho-social skills (self-esteem, self-regulation and self-efficacy), which themselves are challenging outcomes to influence – but noticeably were not different when comparing the SB and control groups in this study. It is also crucial to recognise that the intervention was delivered in a real-world context and that any improvements were measured against the business-as-usual. In this case, the business-as-usual appeared to be a relatively high level of provision, with 90% (19 out of 21) and 71% (15 out of 21) of survey respondents indicating that

in-class and out-of-class adult support, respectively, was available to pupils with communication difficulties. Although this positively reflects the existing teaching provision, it sets a relatively high bar for the SB intervention to surpass, especially as the programme employs one particular approach in supporting a relatively heterogeneous group of pupils.

Finally, the pupil referral criteria included having difficulties with or lacking confidence in communicating, or having poor attention and listening skills. There were no criteria related to factors contributing to the communication challenges. In practice, groups included a mix of children with different needs, which is consistent with previous reports of SB programmes. More than half of the pupils in the sample had EAL. Interviews with school staff indicated that some pupils' communication challenges were related to a special education need or starting school with a lower level of communication skill. Some SB co-facilitators reported that the combination of pupils with different dispositions made it sometimes more difficult to manage the group dynamics of SB sessions, and that pupils' suitability for the programme could be considered both in terms of individual characteristics and group dynamics. It is therefore possible that including a broad mix of children might have limited or delayed the programme's effectiveness at changing the measured outcomes, at least for some pupils.

Overall, the evaluation does not indicate that SB, as delivered in the education context of this study, has a measurable impact on either the primary or secondary outcomes. As such the evaluation does not support SB being delivered more widely in its current form. However, in line with previous evaluations, there is evidence that schools find it a useful tool and that the approach does have the potential to support some pupils. It may therefore be worth revising the intervention and undertaking further research on future iterations of the intervention that are targeted more specifically at pupils with particular difficulties.

## Limitations and lessons learned

### Limitations

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

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

Unfortunately, due to issues in accessing the National Pupil Database, planned pre-test measures (planned to be EYFSP scores) were not available for this evaluation. However, even without these baseline covariates, our MDES at analysis was 0.22 SD, which was only slightly higher than at randomisation (0.21 SD). Due to these data access issues, it was also not possible to obtain the FSM variable indicating whether pupils were ever eligible for FSM in the past 6 years (EVERFSM\_6\_P), and instead we only had access to information indicating whether pupils were eligible for FSM in the school year prior to the intervention. It is therefore possible that the evaluation had missed relevant pupils for this sub-group analysis.

The decision to conduct pupil-level randomisation rested partly on the assumption that the risk of intervention spill-over is low, given that pupils are thought to need multiple SB sessions to experience the intervention's benefits. Nevertheless, there remained a risk that spill-over may have occurred via the TAs, who interacted with pupils from both intervention and control groups in classrooms and other contexts outside the SB sessions. Interviews with TAs indicated that they gained new knowledge, skills and confidence for supporting pupils with speech and language difficulties, and that they understandably implemented some SB strategies outside the sessions. This may have led pupils from the control group to experience some of the intervention's benefits.

An additional concern relates to the inter-rater reliability of the Renfrew Bus Story (RBS) outcome. Although some evidence suggests moderate reliability, others have raised concerns (Hayward et al., 2008). In line with this, we encountered difficulties during the piloting stage with establishing a high inter-rater reliability. We sought to mitigate this by engaging Elklan Training, a speech and language training company with expertise in administering and marking RBS, to score the assessments. Elklan moderated the marking process by blind repeat scoring one in 50 assessments. Nevertheless, it is possible that remaining issues with score reliability may have introduced imprecision in the evaluation.

One limitation in this trial, which is broadly true of most trials, was that the programme was not implemented exactly as it would be in non-trial conditions. Under usual delivery conditions, teachers would be more targeted and selective in their nomination of pupils; whereas for the trial, all schools needed to nominate 40 pupils to be randomised. This may mean the sample was more heterogeneous than under usual delivery in order to meet the referral requirement.

Finally, as discussed above, SB sessions included a mix of children with different needs, in line with the broad pupil referral criteria. The evaluation plan did not include collecting pupil-level data on referral criteria. However, the observed combination of pupil characteristics and underlying causes related to speech and language difficulties may have introduced heterogeneity into the treatment effects. Exploratory sub-group analyses of EAL and non-EAL pupils did not suggest any effects in either sub-group but other unidentified sources of heterogeneity may have remained.

### **Lessons learned**

One significant challenge for this evaluation was the complexity of pupil outcome data collection. We collected four pupil outcome measures, three of which were collected from pupils directly and one from teachers. The pupil assessments took up to an hour to complete –15 minutes to individually administer RBS, and up to 45 minutes to administer the PIRA and WSEM3 in small groups. In an effort to avoid overwhelming pupils, we opted to collect RBS on one day and the PIRA and WSEM3 on a different day. This required two to four visits per school, depending on school schedules, pupil availability and school willingness. Further, we estimated that teachers spent one to two hours each to complete the SSIS survey. Our recommendation would be to carefully consider the amount of data being collected and how much time it will require of schools to support.

Another challenge related to outcome measures is using one that is commonly used by schools. Commonly used assessments carry the benefits of being familiar to schools and having more data on reliability, validity and correlations with other common assessments. However, this can create challenges where the assessment is already being used by participating schools – both for schools who plan to test before or after the evaluator does. It might be worth future evaluators considering data sharing arrangements with schools, particularly where the evaluator anticipates conducting the assessment first.

### **Future research and publication**

The IPE results indicate the importance of further quantitative and in-depth qualitative research of the role and impact of the theorised moderating factors, the types of pupils that might benefit from the intervention, as well as the intermediate outcomes (i.e., self-efficacy and self-regulation). Firstly, more research is needed to understand the ways in which the theorised participant-level (e.g., pupil characteristics and TA engagement) and context-level moderating factors (e.g., parent and teacher involvement, existing support available to pupils with speech and language difficulties) impact outcomes. The IPE findings indicate that some of the theorised moderating factors, such as TAs' knowledge of the pupils and DPs' drama and SB experience, can impact how the programme is implemented. However, it is unclear the extent to which this might impact pupil outcomes. Similarly, further research should explore which pupils might benefit from the intervention, alongside how to best create and manage the group dynamics of pupils with different dispositions (e.g., those who are quiet and shy alongside those who are boisterous). This aligns with the findings from case study interviews with the TAs and DPs, some of whom reported having difficulty managing the behaviours of some pupils who they felt could be disruptive to the session. Research in this area could aid with refining the pupil referral criteria, which is currently relatively broad. Lastly, there is preliminary evidence from the IPE to suggest that the intermediate outcomes of self-efficacy and self-regulation might lead to improvements in written and oral communication skills. Future research, with a longer follow-up period, should test whether and which of these outcomes occur at proximal and distal timepoints after the intervention, as well as whether improvement in these outcomes is correlated.

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

## References

- Adamek, M., Anderson, A., Berry, K.A., Brown, M.R., Darrow, A., Frey, J.R., Friedlaender, L., Glass, D., Kandler, Z., Keddell, K., Kraft, M., Kowitt, J., Lee, B.K., Loesl, S.D., Muller E., Nutting, D., Wenz, C. 2017. *VSA Intersections: Arts and special education*. The John F. Kennedy Center for the Performing Arts.
- 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.
- Barnes, J. 2015. *Speech Bubbles: An evaluation of the 2013–14 extended programme funded by the Shine Trust*. Canterbury Christ Church University.
- Barnes, J. 2018. England: ‘Speech Bubbles’. In Aguiar, C. and Silva, C.S. 2018. Case studies on curriculum, pedagogy, and school climate interventions tackling inequalities (pp. 18–37). *ISOTIS: Inclusive Education and Social Support to Tackle Inequalities in Society*. [online] [http://archive.isotis.org/wp-content/uploads/2019/03/ISOTIS\\_D4.3\\_Case-studies\\_curriculum\\_pedagogy\\_social-climate-interventions.pdf](http://archive.isotis.org/wp-content/uploads/2019/03/ISOTIS_D4.3_Case-studies_curriculum_pedagogy_social-climate-interventions.pdf)
- van Buuren S. and Groothuis-Oudshoorn, K. 2011. Mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3), 1–67.
- Cameron, A.C., Miller, D.L. 2015. A practitioner’s guide to cluster-robust inference. *Journal of Human Resources* 50(2), 317–372.
- Cameron, A.C., Jonah B.G., Miller, D.L. 2008. Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics* 90(3), 414–427.
- Catterall, J.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.S., Dumais, S.A., Hampden-Thompson, G. 2012. *The arts and achievement in at-risk youth: Findings from four longitudinal studies*. National Endowment for the Arts. [online] [https://www.researchgate.net/publication/263479321\\_The\\_Arts\\_and\\_Achievement\\_in\\_At-Risk\\_Youth\\_Findings\\_from\\_Four\\_Longitudinal\\_Studies](https://www.researchgate.net/publication/263479321_The_Arts_and_Achievement_in_At-Risk_Youth_Findings_from_Four_Longitudinal_Studies)
- Coppock, A. 2019. [online] <https://cloud.r-project.org/web/packages/ri2/index.html>.
- Cunningham, S., 2021. *Causal inference: The mixtape*. Yale University Press.
- Dockrell, J.E. 2001. Assessing language skills in preschool children. *Child Psychology and Psychiatry Review* 6(2), 74–85.
- Esarey, J., Menger A 2017. Practical and effective approaches to dealing with clustered data. *Political Science Research and Methods* 1–35 [online], <http://jee3.web.rice.edu/cluster-paper.pdf>.
- Gerber A.S., Green D.P. 2012. *Field experiments*. New York: W. W. Norton & Company.
- Gresham, F.M., Elliott, S.N., 2011. Test review. *Journal of Psychoeducational Assessment* 29(3), 292–296.
- Hannes, K. 2011. Chapter 4: Critical appraisal of qualitative research. In J. Noyes, A. Booth, K. Hannes, A. Harden, J. Harris, S. Lewin, C. Lockwood (Eds.), *Supplementary guidance for inclusion of qualitative research in Cochrane systematic reviews of interventions (version 1)*. Retrieved from [www.cqrmg.cochrane.org/supplemental-handbook-guidance](http://www.cqrmg.cochrane.org/supplemental-handbook-guidance)
- Hayward, D. V., Stewart, G.E., Phillips, L. M., Norris, S. P., Lovell, M. A. 2008. Test review: The Renfrew Bus Story (RBS), in *Language, Phonological Awareness, and Reading Test Directory* (pp. 1–10). Edmonton, AB: Canadian Centre for Research on Literacy. [online] [https://sites.ualberta.ca/~lphillip/documents/AAG-The%20Renfrew%20Bus%20Story%20\(RBS\).doc](https://sites.ualberta.ca/~lphillip/documents/AAG-The%20Renfrew%20Bus%20Story%20(RBS).doc)
- Hetland, L., Winner, E. 2001. The arts and academic achievement: What the evidence shows. *Arts Education Policy Review* 102(5), 3–6.
- Kleiber, C., Zeileis A. 2008. *Applied econometrics with R*. Springer-Verlag, New York. ISBN 978-0-387-77316-2, <https://CRAN.R-project.org/package=AER>.

- Lee, T. 2015. *Princesses, dragons and helicopter stories: Storytelling and story acting in the early years*. Routledge.
- Lee, B.K., Enciso, P., Brown, M.R. 2020. The effect of drama-based pedagogies on K-12 literacy-related outcomes: A meta-analysis of 30 years of research. *International Journal of Education & the Arts* 21(30) [online], <http://doi.org/10.26209/ijea21n30>
- Neelands, J., Belfiore, E., Firth, C., Hart, N., Perrin, L., Brock, S., Holdaway, D., Woddis, J., 2015. *Enriching Britain culture, creativity and growth: 2015 report by the Warwick Commission on the future of cultural value*. University of Warwick, Coventry.
- Newman, M., Bird, K., Tripney, J., Kalra, N., Kwan, I., Bangpan, M., Vigurs, C., 2010. *Understanding the impact of engagement in culture and sport: A systematic review of the learning impacts for young people. CASE: The culture and sport evidence programme*. [online], [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/88447/CASE-systematic-review-July10.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/88447/CASE-systematic-review-July10.pdf)
- Nowell, L.S., Norris, J.M., White, D.E., Moules, N.J. 2017. Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods* doi: 10.1177/1609406917733847
- Podlozny, A. 2000. Strengthening verbal skills through the use of classroom drama: A clear link. *Journal of Aesthetic Education* 34(3/4), 239–275.
- Price, H., Ansong, E. 2016. *'Speech Bubbles' drama intervention programme: Preliminary executive summary of effectiveness*. University of East London.
- Ritchie, J., Lewis, J., Nichols, C.M., Ormston, R., 2014. *Qualitative research in practice*. 2nd edn. London: Sage.
- See, B.H., Kokotsaki, D. 2015. Impact of arts education on the cognitive and non-cognitive outcomes of school-aged children. A review of evidence.
- UK Government, 2019. *Get information about schools*. [online] Available through <https://get-information-schools.service.gov.uk/> [Accessed in February 2019].

## Appendix A: EEF cost rating

Figure A1: Cost rating

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

## Appendix B: Security classification of trial findings

### OUTCOME: PROGRESS IN READING ASSESSMENT

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

Threats to validity	Threat to internal validity?	Comments
<b>Threat 1: Confounding</b>	Low	Randomised design – no apparent confounders.
<b>Threat 2: Concurrent Interventions</b>	Low	It was reported that the schools did take part in lots of other support for children with language difficulties and this was part of their BAU. Thus the current intervention was being compared to a control BAU where pupils were already receiving support. This may make it harder to see effects above and beyond the usual support but this is explicitly discussed in the report, results are presented within this context, and it is measuring the likely real world effect. In addition the trial was pupil randomised suggesting that BAU support offered did not differ across control and intervention groups within schools – there was no mention of the intervention children being refused usual in class support.
<b>Threat 3: Experimental effects</b>	Low	Within school randomisation which can sometimes lead to contamination but this was discussed in the report and deemed that there was unlikely to be spill over to control children as the activities during the session were so bespoke and required the quiet space, attention, and group participation. It was deemed unlikely that these activities would be repeated with the control children.

<b>Threat 4: Implementation fidelity</b>	Low	IPE suggests high fidelity to the intervention. Careful analyses looking at fidelity and CACE compliance. No concerns.
<b>Threat 5: Missing Data</b>	Moderate	Missing data was moderate level but explored using a regression analyses and showed it was missing at random.
<b>Threat 6: Measurement of Outcomes</b>	Low	Outcome measurement well chosen, carefully selected for pupil's ability, and thoroughly implemented. Considerations of school familiarity with the test fully discussed and considered.
<b>Threat 7: Selective reporting</b>	Low	Reporting seems accurate. Analysis plan altered but due to data availability and a new plan was determined before data analysis began and this is explicitly mentioned in the report therefore I have marked this as low.

- **Initial padlock score:** 4 Padlocks
- **Reason for adjustment for threats to validity:** -0 Padlocks – No evidence of serious threats to validity.
- **Final padlock score:** initial score adjusted for threats to validity = 4 Padlocks – no adjustment needed.

#### OUTCOME: RENFREW BUS STORY

Please use this template to assign a separate security rating for each primary outcome.

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

Threats to validity	Threat to internal validity?	Comments
<b>Threat 1: Confounding</b>	Low	Randomised design – no apparent confounders.
<b>Threat 2: Concurrent Interventions</b>	Low	It was reported that the schools did take part in lots of other support for children with language difficulties and this was part of their BAU. Thus the current intervention was being compared to a control BAU where pupils were already receiving support. This may make it harder to see effects above and beyond the usual support but this is explicitly discussed in the report, results are presented within this context, and it is measuring the likely real world effect. In addition the trial was pupil randomised suggesting that BAU support offered did not differ across control and intervention groups within schools – there was no mention of the intervention children being refused usual in class support.
<b>Threat 3: Experimental effects</b>	Low	Within school randomisation which can sometimes lead to contamination but this was discussed in the report and deemed that there was unlikely to be spill over to control children as the activities during the session were so bespoke and required the quiet space, attention, and group participation. It was deemed unlikely that these activities would be repeated with the control children.
<b>Threat 4: Implementation fidelity</b>	Low	IPE suggests high fidelity to the intervention. Careful analyses looking at fidelity and CACE compliance. No concerns.
<b>Threat 5: Missing Data</b>	Moderate	Missing data was moderate level but explored using a regression analyses and showed it was missing at random.
<b>Threat 6: Measurement of Outcomes</b>	Low	Outcome measurement well chosen, piloted to check suitability, and thoroughly implemented. Extra care taken with scoring. Limitations fully explored.
<b>Threat 7: Selective reporting</b>	Low	Reporting seems accurate. Analysis plan altered but due to data availability and a new plan was determined before data analysis began and this is explicitly mentioned in the report therefore I have marked this as low.

- **Initial padlock score:** 4 Padlocks
- **Reason for adjustment for threats to validity:** -0 Padlocks – No evidence of serious threats to validity.
- **Final padlock score:** initial score adjusted for threats to validity = 4 Padlocks – no adjustment needed.

## Appendix C: Effect size estimation

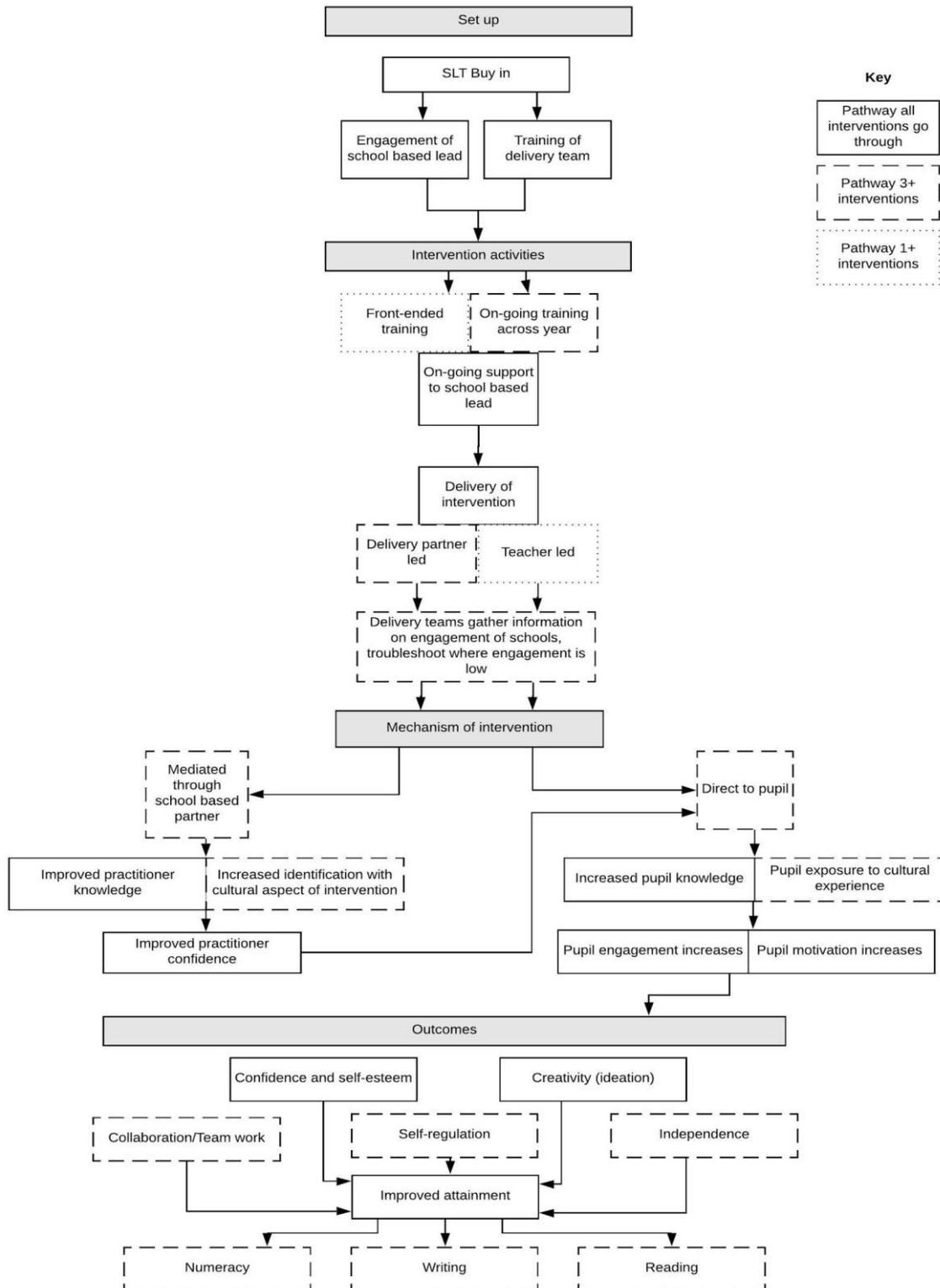
Table C1: Effect size estimation

Outcome	Unadjusted differences in means	Adjusted differences in means	Intervention group		Control group		Pooled SD	Hedges' correction (J)
			<i>n</i> (missing)	SD of outcome	<i>n</i> (missing)	SD of outcome		
PIRA	-0.07	-0.05	411 (91)	0.99	410 (94)	1	1	0.99908
RBS	-0.35	-0.46	405 (97)	11.09	406 (98)	11.52	11.31	0.99907
WSEM3	0.07	0.08	377 (125)	1.44	393 (111)	1.5	1.47	0.99902
SSIS	0.51	0.63	383 (119)	23.07	363 (141)	21.8	22.46	0.99899

Note. SD refers to standard deviation.

## Further Appendices

### Appendix D: Overarching flow diagram of programme similarities



## Appendix E: Speech Bubbles referral guidance

Dear Teacher,

Speech Bubbles is a KS1 intervention using drama to help children develop their communication skills. The intervention is run by an external drama practitioner and a member of school support staff.

Speech Bubble sessions run 8 weeks per term, either side of half term. 20 KS1 children will be mixed into two groups of ten. Sessions last 40 minutes.

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.

If you have any questions about the project throughout the year please ask the staff involved. If you have time to visit a session you are very welcome.

## Appendix F: Speech Bubbles CPD induction day details

DATE OF SUBMISSION:	30 <sup>th</sup> June 2014
TITLE OF CPD COURSE:	The Speech Bubbles approach. Drama for Communication and Personal & Social Wellbeing.
BRIEF DESCRIPTION:	A practical workshop that introduces the ethos, practice and evidence of the Speech Bubbles project.  A follow up workshop that gives participants the opportunity to reflect on the experience of delivering Speech Bubbles and gain additional practical knowledge.
WEBSITE ADDRESS:	<a href="http://www.londonbubble.org.uk">www.londonbubble.org.uk</a>
TARGET AUDIENCE(S):	Drama practitioners, teachers, school support staff, educational psychologists, Speech and language Therapists that work with children aged 5-7.
ANTICIPATED DURATION: (No. CPD Hours)	6 ¼ Hours
TYPE OF CPD: (Seminar, training course, event, conference, distance learning, online learning, etc.)	Practical workshop and follow up workshop.  Resource materials
SUMMARY OF CONTENT:	<ul style="list-style-type: none"> <li>● Introduction to planning, delivery and evaluation of small group drama for communication.</li> <li>● Practical exploration of proven drama techniques that enrich, develop and support children's communication skills, confidence and self esteem.</li> <li>● Awareness of Speech, Language and communication needs.</li> <li>● A resource pack including sections on planning, workshop structure, resources &amp; activities</li> </ul>

<p>LEARNING OBJECTIVES:</p>	<ol style="list-style-type: none"> <li>1. Participants to gain an overview of the Speech Bubbles project/approach</li> <li>2. Participants to have increased awareness of Speech Language and Communication Needs in children aged 5-7               <ul style="list-style-type: none"> <li>○ Evidence of need</li> <li>○ Evidence of impact</li> <li>○ Connection with social and personal wellbeing</li> </ul> </li> <li>3. Participants to have tried out practical drama activities that promote wellbeing and communication - considering               <ul style="list-style-type: none"> <li>○ Group ethos</li> <li>○ Planning</li> <li>○ Evaluation</li> </ul> </li> </ol>	
<p>ESTIMATED LEVEL OF AUDIENCE KNOWLEDGE:  (PLEASE TICK)</p>	<p>ENTRY/BEGINNER LEVEL: – material is introductory and designed to impart basic awareness of the topic. Fundamentals and key issues will be addressed</p>	
	<p>INTERMEDIATE LEVEL: – material is appropriate for all levels. Novices and experienced participants alike will benefit, gaining general understanding and comprehension of the material</p>	X
	<p>ADVANCED LEVEL: – material is advanced and geared for those with a firm grasp of the subject. Participants should have prior experience to apply this material to existing skills</p>	

## Appendix G: School information sheet and MOU



### Learning About Culture - Randomised Control Trial of Speech Bubbles

London Bubble Theatre are looking for 10 South London schools to take part in a research trial of the **Speech Bubbles** drama for communication intervention. This is one of **five trials** taking place in 2018-19. 9000 children in 400 schools across England will be studied to find out if different cultural learning approaches can help boost primary pupils' achievement.

Our trial will evaluate the effect of Speech Bubbles on the communication skills and reading abilities of KS1 children with SLCN. The impact on pupil engagement will also be measured. The results of this project will make a valuable contribution to understanding the value of drama and storytelling in improving pupil attainment.

In recognition of the commitments involved research schools will receive the full Speech Bubbles intervention at the greatly reduced price of **£500+VAT**.

#### Who is involved?



Theatre Company who lead the delivery of Speech Bubbles. Provide all training, resources and theatre practitioners to deliver the programme in schools.



Co-lead partner, designing the research programme and disseminating results.



Co-lead partner, designing the research programme and disseminating results.



Researchers working on collection and processing of data.



Researchers working on collection and processing of data.

#### Who is Speech Bubbles For?

**Children who will be in KS1 in 2018/19, 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.

For more information:  
Visit [www.speechbubbles.org.uk](http://www.speechbubbles.org.uk)  
or contact Speech Bubbles  
Coordinator Amelia Bird  
020 7237 4434  
[amelia@londonbubble.org.uk](mailto:amelia@londonbubble.org.uk)



### **Research Project Timeline and Commitments**

#### **DECEMBER 2017-MARCH 2018: SCHOOLS AGREED**

Schools must have a minimum of 2 form entry and not have had Speech Bubbles since 2013/14.

#### **EARLY MAY 2018: 40 APPROPRIATE RECEPTION/YEAR1 CHILDREN REFERRED**

Opt-out consent forms sent home.

#### **LATE MAY 2018: RESEARCH CHILDREN CONFIRMED AND BASIC DATA SENT**

Children who have not opted out confirmed as the 'research children'. Basic data on pupil premium, EAL, gender, age etc. sent to evaluators.

#### **JUNE 2018: BASELINE TESTS**

PIRA reading test led by teacher on all research children. (TBC)

#### **LATE JUNE 2018: SCHOOLS FULLY CONFIRMED**

Schools must have completed protocols and baseline tests to progress on to project.

#### **EARLY JULY 2018: RANDOMISATION**

Researchers assign 20 children to Speech [Bubbles](#), the remaining research children form the control group.

#### **SEPTEMBER 2018: SPEECH BUBBLES INDUCTION**

School staff complete Speech Bubbles online assessment for 20 children. 2 members of staff attend full day of CPD certified training.

#### **SEPTEMBER 2018-JULY 2019: SPEECH BUBBLES**

20 children take part in the Speech Bubbles programme. Designated member of staff (usually a TA) released for half a day per week to work on the programme.

#### **JANUARY 2019: COLLABORATIVE LEARNING**

Designated member of staff attends half day CPD certified learning session.

#### **APRIL-JUNE 2019: QUALITATIVE EVALUATION**

Questionnaire completed by teachers and teaching assistants. A handful of schools interviewed on their experiences of Speech Bubbles.

#### **MAY-JULY 2019: POST ASSESSMENTS**

Speech Bubbles assessment completed by class teacher for 20 children.  
PIRA Reading test led by Research Assistant on all research children.  
Communication Assessment led by a Research Assistant on all research children.  
SSIS Social Skills Measure questionnaires completed by teachers on all research children.

#### **EARLY JULY 2019: EVALUATION SESSION**

Designated member of staff attends standard half day Speech Bubbles evaluation.

#### **SPRING/SUMMER 2020: FINAL RESEARCH REPORT LAUNCHED**

## **Aims of the evaluation**

The aim of this project is to evaluate the effect of Speech Bubbles - a creative drama programme - on the reading, communication and social skills of Year 1 and 2 children with speech, language and communication needs. The results of this project will make an important contribution to understanding the value of drama and storytelling in improving pupil attainment.

## **The project**

Speech Bubbles will be delivered by London Bubble and [partner organisation], and is funded by the Education Endowment Foundation (EEF). Participating schools will be asked to refer children who will be moving into Year 1 and 2 in 2018/19 in accordance with the referral guidance and to gain opt-out consent from the parents/carers of these children in the 2017/18 school year. The impact of three terms of Speech Bubbles will be evaluated using a randomised controlled trial (RCT).

During this project, you will be contacted by both London Bubble and [partner organisation], hereafter referred to as the 'Project Team', who are responsible for Speech Bubbles delivery, support and resources, and by the Behavioural Insights Team, UCL Institute of Education and the Royal Society of Arts (RSA), hereafter referred to as the 'Evaluation Team', who are carrying out an independent evaluation of project.

## **Structure of the evaluation**

Children who are referred into the programme will have a 50 per cent chance of being selected to take part. The remaining children will form the control group and will not be offered a place in Speech Bubbles classes. Children who take part will participate in 24, 45-minute weekly drama sessions during the 2018/19 school year.

Random assignment of children to the control or treatment arms is essential to the evaluation. It is important that schools understand that children will be assigned to either group and that they are agreeing to abide by this assignment.

Reading attainment is the primary outcome of interest in this evaluation, and will be measured using the Progress in Reading Assessment (PIRA) by Rising Stars. Research Assistants (RAs) employed by the Evaluation Team will administer PIRA as well as the Renfrew Bus Story test (a test of communication skill that is conducted 1:1) in summer 2019. A survey measuring child social skill and cultural engagement will also need to be completed by classroom teachers at this time.

## **Use of Data**

All data, including pupils' test responses and any other pupil data, will be treated with the strictest confidence. Pupil assessments will be marked by RAs hired by the Evaluation Team, or the test publisher. All sensitive data will be stored securely by the Evaluation Team. Named data will be matched with the National Pupil Database and shared with UCL, the Department for Education, the EEF, FFT Education, and in an anonymised form the UK Data Archive. All results will be anonymised so that no schools or individual pupils will be identifiable in the report or dissemination of any results. Confidentiality will be maintained and no one outside the Project Team and Evaluation Teams will have access to the database. The only exception to this is if a child tells us something that raises concerns regarding child safeguarding/wellbeing when we are obliged to inform your safeguarding officer and liaise with them on appropriate action.

## **Responsibilities**

### **The London Bubble Theatre Company agrees:**

- To devise and implement drama workshops as agreed for referred young people.
- To plan and implement training and evaluation sessions for drama practitioners and school staff.
- To put in place a suitable contingency if drama practitioner cannot attend a session i.e. sickness. (Acceptable contingencies: rescheduling of session by agreement with school or providing cover practitioners).
- Work closely with the school to encourage parental involvement and attendance.
- To ensure drama practitioners have a current DBS enhanced disclosure.
- To report any child protection concerns to the school's designated person.
- To invoice the agreed fee of £500 plus VAT.

**The Behavioural Insights Team, UCL Institute of Education and The Royal Society of Arts agree:**

- To conduct the random allocation of schools/children to trial arms.
- To collect class and pupil level data; including UPNs, first and last names, DoB, FSM status, year level, EAL status and gender for all referred children (barring those who were opted out)
- To collect assessments of aforementioned outcome measures.
- To store and transfer all data safely and securely.
- To analyse data from the project in order to produce impact estimates.
- To conduct the implementation and process evaluation, including analysis and reporting from this.
- To produce end of project evaluation report to be published by the Education Endowment Foundation.

**Primary School agrees:**

- To participate in the trial, including the collection of outcome measures.
- To enable the Speech Bubbles sessions to take place in a space suitable for practical small group drama within the times discussed and agreed.
- To allocate a member of the school support staff to attend the Induction day and support the weekly delivery of the programme.
- To create a suitable contingency if allocated support staff member cannot attend session i.e. sickness. (Acceptable contingencies: provide an alternative member of staff or rescheduling of session by agreement with London Bubble Theatre Company)
- To complete pre-project teacher assessments before sessions commence and post-project assessments at the end of the year.
- To have gained agreement from 2018/19 Year 1 and 2 teachers and teaching assistants for collaborating with the Evaluation Team when required, chiefly with respect to completing surveys concerning the social skills and cultural engagement of children in the study, and allowing reading and communication assessments to be collected during class time.
- To make Year 1 and Year 2 staff available for the purposes of interviews in order to complete a qualitative assessment of the programme, and to complete a brief survey about their involvement in the project.
- To refer children to the project according to the referral guidance, distribute opt-out consent forms to parents of these children, and to keep a record of which pupils are opted-out.
- For those children whose parents do not opt out, to provide the Evaluation Team with the required pupil data.
- To abide by the random assignment of children to the Speech Bubbles programme, and NOT offer this programme to any children allocated to the control condition.
- To inform the Evaluation Team in the event any participating children cease to participate in the Speech Bubbles intervention or leave the school (including for children in the control condition).
- To nominate a staff member to act as a primary contact, which involves attending planning and progress meetings in order to ensure the success of the project, taking responsibility for the timely distribution of consent forms to parents, provision of pupil data records, coordination of summer 2018 reading assessments and scheduling of any Evaluation or Project Team staff visits for the purpose of conducting assessments or performing qualitative research.

**Contact Details**

Thank you for agreeing to take part in this research. Please do not hesitate to get in touch if you have any questions about the Speech Bubbles programme or the evaluation itself.

[CONTACT DETAILS REDACTED]

**Agreement**

I agree for my school to take part in the evaluation of Speech Bubbles research project and I accept the requirements outlined in this MoU.

## Appendix H: Information sheet for parents

Dear Parent / Carer,

Your child's school is taking part in a research programme that aims to improve reading, communication and social skills. The Speech Bubbles programme, run by London Bubble and [partner organisation], was developed by theatre artists, speech and language therapists, and schools to promote children's communication, confidence and wellbeing. The approach encourages children to tell, act out and reflect on their own stories. Speech Bubbles staff create a safe, playful and fun environment, which means that children participating in the programme feel they are having a weekly treat.

Your child has been referred for participation in Speech Bubbles by their teacher, but as this is a research project, not all children referred will actually take part. Whether your child attends Speech Bubbles classes will be decided randomly by evaluators (The Behavioural Insights Team, UCL Institute of Education and The Royal Society of Arts) in order to look at how effective Speech Bubbles has been.

As part of measuring the success of Speech Bubbles, children will complete a reading assessment in summer 2018 and 2019, and a language assessment in summer 2019. Your child's teacher will conduct the reading assessment in 2018, and staff hired and trained by the evaluator will conduct both the reading and language assessments in 2019. At this time a survey concerning your child's social skills will also be completed by their teacher. These assessments will not be used to monitor teacher or school performance, and are only collected to help us understand how much Speech Bubbles benefits children.

Your child's name and other data held by the school, alongside their reading assessment scores, will be collected by the evaluators, in conjunction with London Bubble and contracted test markers. No information that can identify individual children will be made available to anyone outside these teams and your child's school. The only exception to this is if your child tells us something that raises concerns regarding child safeguarding/wellbeing at which point we will inform the school's safeguarding officer and liaise with them on appropriate action. The data will be kept under password protection in an offline format. We will not use your child's name or the name of the school in any report arising from the research, and no information that could otherwise identify your child will be made public.

We will also obtain your child's UPN (Unique Pupil Number) to allow longer term understanding of whether Speech Bubbles is effective. We will save this information in a data format that will prevent anyone from identifying your child. This data will then be linked with the National Pupil Database (held by the Department for Education), other official records, and shared with the project team, the Department for Education, Education Endowment Foundation (EEF), EEF's data contractor FFT Education and in an anonymised form to the UK Data Archive.

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

If you have any questions you would like to ask, please contact Louise Jones on 0780 4494899 or via email at [louise.jones@bi.team](mailto:louise.jones@bi.team).

**If you are happy for information about your child to be used in the Speech Bubbles research project you do not need to do anything. Thank you for your help with this research, your support is much appreciated.**

If you **DO NOT** want information about your child to be used to understand whether the Speech Bubbles programme can help children to improve their reading, communication and social skills, please complete the

enclosed form and return it to your child's school by [INSERT DATE]. If you do this then no information about your child will be shared with the evaluation or project teams at any point during the project.

**With thanks from the Research and Evaluation Teams**

**Speech Bubbles research programme**

(If you are happy for your child to participate in the research on whether this programme improves reading, communication and social skills, you **DO NOT** need to return this form.)

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

Child's name: .....Date of birth: .....

Child's class Teacher: .....

School:.....

Parent name (BLOCK CAPITALS) .....

Parent signature: .....

Date .....

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

## Appendix I: Piloting the Renfrew Bus Story test

### Sample

- A total of 88 children were tested across 3 schools in London and Manchester, 48 boys and 40 girls
- Children were between 5 and 7 years old, with 6 year olds making up 41% of the sample
- 78 children were part of the Speech Bubbles Programme (89%)
- 44% of children had EAL
- 20% had SEN

### Overall engagement

- Children engaged with the test very well, including children with selective mutism
- Those who were anxious or shy to start with, became comfortable fairly quickly
- Two children did not engage at all, although they were not distressed or upset
- Overall, the majority of children displayed partial to full engagement with the Bus Story test.

Table 1. Spectrum of pupil engagement

Spectrum of Engagement				
Low		→		High
	No response to story	Looking up/around the room	Reciting story with researcher's prompts	Reciting story without any prompts
		Looking/smiling at the teacher	Mostly looking/pointing at picture book when reciting	Turning picture book/pointing themselves
		Looking/smiling at other children (if present)	Reciting story with minor notice of the audio recorder	
<b>Proportion of children (approx.)</b>	5%	15%	55%	25%

### Recommendations for improving children's engagement

- Encourage children to "concentrate on the book" and "do their best" before starting to the test
- Verbalise consistent encouragement throughout i.e "well done", "keep going" and "that's great"
- Use non-verbal cues such as nods and pointing to the next picture
- Use the above cues as soon as children seem unsure, distracted, or look up around the room, at the researcher or at others
- If possible, have the audio recorder out of sight

### Recommendations: Research Assistant Training

- Make training as interactive as possible. It should include some demonstrations from Lauren or other experienced administrator, as well as role playing

- Emphasise the value in using a calm demeanour with students
- Ideally, research assistants will have experience in early years setting
- Carefully train research assistants on ethical responsibilities and how to recognise signs of distress in young children:
  - Students may opt out by refusing to come to the assessment space (but should be offered another opportunity later in the day if possible)
  - Any prompting should adhere to the administration script, which is not coercive
  - Administration should end if student appears distressed

## Teacher Presence

- Teachers collected children from their classrooms and walked them to the testing room:
  - This calmed children's nerves and acted as a helpful introduction between the researcher and the child;
  - Children appeared particularly calm with teachers they were familiar with i.e welfare officers and Speech Bubbles tutors.
- Having a familiar adult in the room (TA, welfare officer, Speech Bubbles tutor):
  - The adults were able to detect when the child was distressed and assist.
- Children were comfortable reciting the story with and without a TA present
- The teachers varied in their support/involvement and offering of prompts:
  - Some teachers intervened by reassuring children who were nervous, without directly offering information about the story which was helpful;
  - Other teachers would intervene when it wasn't necessary, offering prompts that were not recommended and likely influenced children's performance;
  - Some teachers gave prompts as the researcher was about to, which caused a little confusion/disruption for the children

## The environment

- On occasion, children who were waiting to be tested would do so in the testing room while drawing/reading:
  - Collecting multiple children at a time reduced classroom disruption;
  - If the child testing was in view of the other children they would make eye contact, smile, and lose concentration;
  - This was not an issue if the children were out of sight of the child who was assessed;
  - Children who heard the story repeatedly might have had an unfair advantage.
- Shared spaces i.e. school canteens, were subject to a lot of noise/distraction:
  - This negatively affected children's concentration;
  - Private classrooms allocated to Bus Story testing were quieter and provided a better testing environment.

Table 2. Spectrum of environment distraction

Environment Distraction		
Low	→	High
Private in-school classroom	Private classroom lodge located in the playground	School canteen next to playground
Quiet room away from corridor and outside noise/distraction	Others couldn't enter the room but playground noise was loud/distraction during break times	Communal area where teachers would enter/leave  Playground noise/distraction

### Recommendations: teacher presence

- Teachers to collect children from their classrooms, and briefly explain that they will be listening to a story about a bus (not referring to it as a 'test')
- Teachers to bring children to the testing room and introduce them to the researcher;
- Teachers should not be present during the test to eliminate variation in additional support which might influence children's performance.<sup>63</sup> But they should ideally be nearby in case additional social-emotional support is needed.

### Recommendations: school coordination

- Provide teachers with a one-pager in advance that contains a list of requirements for testing, as well as considerations
- Clearly communicate that all children should have the opportunity to participate; teachers should not be deciding whether students are capable of completing the assessment. But also ensure teachers know students should not be forced to participate if they do not want to come; can try again later in the day or mark the student as having "opted out"
- Requirements to include:
  - space to conduct test (ideally, quiet and with minimal distractions)
  - assistance with pulling students who will participate in testing (ideally, one at a time)
- Considerations to include:
  - appropriate lesson plan where kids coming in and out will not be distracting or too detrimental to learning objective
  - any 1:1 support student might need to engage with assessment

### Recommendations: test delivery

- Test setup:
  - Private room free from distractions/noise.

<sup>63</sup> For students who typically have 1:1 support during the school day, a support person may be present for social-emotional support only.

- Establish some way of coding quality of the testing environment
- Researcher should:
  - Introduce themselves to each child and explain the task;
  - Check the child can see the picture book and understands the task;
  - Speak at a clear and a slow pace;
  - Offer story starting prompt even if child says they don't want to tell the story;
  - Allow time before offering prompts.

### Results: Marking

- The quietest children can be heard during playback of the audio recordings
- All transcription was completed by research assistants
- The marking is challenging and research assistants conducting the assessment will require training from a speech and language specialist who has experience with the Renfrew Bus Story

### Recommendations: Marking

- Recordings should be transcribed directly into scoring forms (see Table 3 below)
- A speech and language therapist or trained research assistant will be required to mark the recordings
- Develop an appropriate data protection plan for the recordings and transcriptions

Table 3. Example marking table

ID	Information Score	Transcription	Sentence Length	Subordinate Clauses
		Once upon a time there was a very naughty bus		
		The bus ran away		
		The bus driver tried to catch him		

### Key recommendations

1. The test is suitable to deliver in schools to the vast majority of pupils
2. Familiar teacher to collect child from classroom, reassure them and introduce them to the researcher
3. Researcher to provide verbal and non-verbal encouragement throughout the test
4. To have no teacher or other children present during the test
5. To test in a private room free from distractions
6. A speech and language therapist to mark all tests (or to train RAs in marking)

## Appendix J: Adapted ideation subscale of the WSEM



Name: \_\_\_\_\_

Instructions: Put a tick in one box for each statement below.

	I don't do this much	I do this sometimes	I do this lots
1. I think of new ideas.			
2. I tell my ideas to other people.			
3. I do things with my ideas.			

Thank you for helping us by thinking about your ideas.

## Appendix K: Randomisation code

```

global dta PATH_TO_DATA
cd OUTPUT_DIR_PATH

log using filename

use "${dta}data for randomisation", clear

tab school

//randomise
set seed 06121979 //set seed
bysort school year : gen rand = runiform() //random var
bysort school year : gen even = !mod(_N, 2) //indicate which stratas have an even number of students
bysort school year : gen n_n //count students in each strata
bysort school year : gen upto = cond(even==1, _N, _N-1)

tab school year, sum(even) //check
bysort school year (rand) : egen treat1 = seq() if n<=upto, f(0) t(1) //randomise upto the even number
tab treat1

egen treat2 = seq() if n>upto, f(0) t(1) //randomise the remaining observations
tab treat2

egen treat = rowtotal(treat*)

drop treat1 treat2 upto even rand n

lab de treat 0 "Control" 1 "Treatment"
lab val treat treat

//balance checks
tab treat
prtest year, by(treat)
prtest gender, by(treat)
prtest fsm, by(treat)
prtest eal, by(treat)

//labels
lab var urn "URN"
lab var school "School"
lab var year "Year Group"
lab var gender "Gender"
lab var fsm "FSM"
lab var eal "EAL"
lab var BITID "BIT ID"
lab var treat "Treatment Assignment"

save "${dta}randomisation"

merge 1:1 upn surname using "${dta}randomisation"

list school upn _merge if _merge!=3

//generate a spreadsheet for each school with children who will take part
cd "${output}"
keep school upn surname firstname year treat

levelsof school, l(schools)
foreach school of local schools {
    preserve
        keep if school=="`school'"
        drop school
        sort treat year

        export excel "`school'", first(var1)
    restore
}

```

## Appendix L: Results of analysis to predict missingness in outcome data

Table L1: Predicting missingness in outcome data – logistic regression analysis results

Coefficient name	Coefficient p-value	
	PIRA outcome missingness	RBS outcome missingness

Treatment	0.881	0.988
FSM	0.662	0.407
EAL	0.139	0.801
Year	0.814	0.539
School 2	0.683	0.395
School 3	0.611	0.904
School 4	0.593	0.818
School 5	0.647	0.553
School 6	0.478	0.567
School 7	0.869	0.836
School 8	0.448	0.784
School 9	0.508	0.688
School 10	0.975	0.975
School 11	0.657	0.944
School 12	0.368	0.447
School 13	0.003	0.019
School 14	0.925	0.665
School 15	0.146	0.914
School 16	0.098	0.130
School 17	0.211	0.249
School 18	0.002	0.000
School 19	0.881	0.348
School 20	0.367	0.415
School 21	0.994	0.821
School 22	0.489	0.256
School 23	0.652	0.397

Speech Bubbles  
Evaluation Report

School 24	0.099	0.124
School 25	0.340	0.394
School 26	0.866	0.425

## Appendix M: Analysis syntax

```
#-----SETUP-----

# load packages
packages <-
c("tidyverse", "data.table", "lme4", "broom", "AER", "clusterSEs", "readstatal3", "pwr", "ri2",
"readxl", "ggpubr")
lapply(packages, library, character.only = TRUE)

# RNG seed
set.seed(20102020)

# number of repetitions of cluster bootstrapping
nbootreps <- 5000

# load data
data <- as.data.table(read.csv(PATH_TO_DATA))

outputDir <- PATH_TO_OUTPUT_DIR

#-----PREP DATA-----

# make factors
cols <- c("year", "fsm", "eal", "treat")
data[, (cols) := lapply(.SD, function(x) relevel(factor(x), ref = "0")), .SDcols =
cols]

# isComplier = as.numeric(isComplier),
data[, ':='(anonschoolid = factor(anonschoolid))]

# fix compliance flag for one school (has compliance data but flag not set!)
data[anonschoolid==10, isComplier := ifelse(sessAttended<16, FALSE, TRUE)]

# set compliance for control group pupils with sessAttended < 16 to FALSE (initially
set this to TRUE)
# compliance = 0 for all controls with sessAttended < 16, = 1 for treated or control
with sessAttended >= 16 (1 pupil)
data[treat==0 & !is.na(isComplier), isComplier := ifelse(sessAttended<16, FALSE, TRUE)]

# as per SAP, normalise PIRA scores within each year to have mean = 0, sd = 1
data[!is.na(piraScore), piraScoreNorm := (piraScore - mean(piraScore, na.rm = TRUE)) /
sd(piraScore, na.rm = TRUE), by = year]

# check
data[!is.na(piraScoreNorm), .(meanOrig = mean(piraScore),
sdOrig = sd(piraScore),
meanNew = mean(piraScoreNorm),
sdNew = sd(piraScoreNorm)), by = year]

#-----MDES TABLE-----

# PROTOCOL STAGE
tableMDES_protocol <- setDT(data.frame(outcome = c("pira", "rbs", "pira_fsm", "rbs_fsm"),
n_schools = c(23, 23, 23, 23),
n_treat = c(460, 460, 65, 65),
n_control = c(460, 460, 65, 65),
n_total = c(920, 920, 130, 130)))

attrition <- 0.2
testretest_corr <- 0.3
```

```

tableMDES_protocol[, MDES := sapply(seq_len(.N),
                                   function(x) {round(pwr.t2n.test(n1 = n_treat[[x]] -
n_treat[[x]]*attrition,
                                   n2 = n_control[[x]]
- n_treat[[x]]*attrition,
                                   power = 0.8,
                                   sig.level =
0.05)$d*sqrt(1-testretest_corr^2),2)}}]
sqrt(1-testretest_corr^2), 1),2)}}]

# transpose and output
tableMDES_protocol <- as.data.frame(t(tableMDES_protocol))

# RANDOMISATION STAGE
tableMDES_rand <- setDT(data.frame(outcome = c("pira","rbs","pira_fsm","rbs_fsm"),
                                   n_schools = c(rep(data[,
length(unique(anonschoolid))],4)),
                                   n_treat = c(data[treat==1, .N],
                                               data[treat==1, .N],
                                               data[treat==1 & fsm==1, .N],
                                               data[treat==1 & fsm==1, .N]),
                                   n_control = c(data[treat==0, .N],
                                                data[treat==0, .N],
                                                data[treat==0 & fsm==1, .N],
                                                data[treat==0 & fsm==1, .N]),
                                   n_total = c(data[, .N],
                                               data[, .N],
                                               data[fsm==1, .N],
                                               data[fsm==1, .N]))

attrition <- 0.2
testretest_corr <- 0.3
tableMDES_rand[, MDES := sapply(seq_len(.N),
                                function(x) {round(pwr.t2n.test(n1 = n_treat[[x]] -
n_treat[[x]]*attrition,
                                n2 = n_control[[x]] -
n_treat[[x]]*attrition,
                                power = 0.8,
                                sig.level =
ifelse(outcome[[x]] %in% c("pira","rbs"),0.05/2,0.05))$d
                                * sqrt(1-testretest_corr^2),2)}}]

# transpose and output
tableMDES_rand <- as.data.frame(t(tableMDES_rand))

# ANALYSIS STAGE
tableMDES_analysis <- setDT(data.frame(outcome = c("pira","rbs","pira_fsm","rbs_fsm"),
                                       n_schools = c(data[!is.na(piraScoreNorm),
length(unique(anonschoolid))],
                                       data[!is.na(bsScore),
length(unique(anonschoolid))],
                                       data[!is.na(piraScoreNorm) &
fsm==1, length(unique(anonschoolid))],
                                       data[!is.na(bsScore) & fsm==1,
length(unique(anonschoolid))]),
                                       n_treat = c(data[!is.na(piraScoreNorm) &
treat==1, .N],
                                       data[!is.na(bsScore) & treat==1,
.N],

```

```

treat==1 & fsm==1, .N],
fsm==1, .N]),
treat==0, .N],
.N],
treat==0 & fsm==1, .N],
fsm==1, .N]),
.N],
.N])))

n_control = c(data[!is.na(piraScoreNorm) &
data[!is.na(bsScore) & treat==1 &
data[!is.na(piraScoreNorm) &
data[!is.na(bsScore) & treat==0,
data[!is.na(piraScoreNorm) &
data[!is.na(bsScore) & treat==0 &

n_total = c(data[!is.na(piraScoreNorm), .N],
data[!is.na(bsScore), .N],
data[!is.na(piraScoreNorm) & fsm==1,
data[!is.na(bsScore) & fsm==1,

tableMDES_analysis[, MDES := sapply(seq_len(.N),
function(x) {round(pwr.t2n.test(n1 = n_treat[[x]],
n2 =
power = 0.8,
sig.level =

ifelse(outcome[[x]] %in% c("pira", "rbs"), 0.05/2, 0.05)}$d, 2)}))

# transpose and output
tableMDES_analysis <- as.data.frame(t(tableMDES_analysis))

#-----GET NATIONAL DATA-----
# spine data (school type)
dataSpine <- suppressWarnings(as.data.table(read_xlsx(PATH_TO_ENGLAND_SPINE_DATA)))

dataSpine <- dataSpine[!duplicated(dataSpine) & !is.na(URN),
.(URN = as.character(URN),
NFTYPE,
schooltype = case_when(NFTYPE %in% c("Academy - Converter
mainstream",
"Academy sponsor led
mainstream",
"Free school -
Mainstream") ~ "academy",
NFTYPE=="Community school" ~
"community",
NFTYPE %in% c("Foundation school",
"Voluntary aided school",
"Voluntary controlled
school") ~ "other"))]

# KS2 data (KS1 scores, FSM, EAL)
dataKS2 <- as.data.table(read.csv(PATH_TO_COMPARE_SCHOOL_PERFORMANCE_DATA))

dataKS2 <- dataKS2[!duplicated(dataKS2) & !is.na(URN), .(URN = as.character(URN),
TKS1AVERAGE =
as.numeric(as.character(TKS1AVERAGE)),
PTFSM6CLA1A =
parse_number(as.character(PTFSM6CLA1A)),
PTEALGRP2 =
parse_number(as.character(PTEALGRP2)))]

# Ofsted ratings

```

```

dataOfsted <- as.data.table(read_xlsx(PATH_TO_OFSTED_DATA))

dataOfsted <- dataOfsted[!duplicated(dataOfsted) & !is.na(URN), .(URN =
as.character(URN), `Overall effectiveness`)]

dataOfsted[, `Overall effectiveness`:= case_when(`Overall effectiveness`==1 ~
"outstanding",
                                                `Overall effectiveness`==2 ~ "good",
                                                `Overall effectiveness`==3 ~
"requiresImprovement",
                                                `Overall effectiveness`==4 ~
"inadequate")]

# edubase (UrbanRural)
dataEdubase <- as.data.table(read.csv(PATH_TO_GIAS_DATA))
dataEdubase <- dataEdubase[!duplicated(dataEdubase) & !is.na(URN), .(URN =
as.character(URN), UrbanRural..name.)]

# merge all national data
dataNational <- Reduce(function(x,y) merge(x, y, by = "URN", all.x = TRUE),
list(dataKS2,dataSpine,dataOfsted,dataEdubase))

# exclude school types
dataNational <- dataNational[!NFTYPE %in% c("Other independent school",
"Community special school",
"Academy - Converter special school",
"Academy - Sponsor led special school",
"Foundation special school",
"Free school - Special")]

dataNational <- dataNational[!is.na(URN) & !is.na(NFTYPE) & !duplicated(dataNational)]

# get national level figures
nationalFigures <- dataNational[, .(propUrban =
round(sum(grepl("Urban",UrbanRural..name., ignore.case = TRUE))/N*100,1),
propRural =
round(sum(!grepl("Urban",UrbanRural..name., ignore.case = TRUE))/N*100,1),

propAcademy =
round(sum(schooltype=="academy")/N*100,1),
propCommunity =
round(sum(schooltype=="community")/N*100,1),
propOther =
round(sum(schooltype=="other")/N*100,1),

propOutstanding = round(sum(`Overall
effectiveness`=="outstanding", na.rm = TRUE)/N*100,1),
propGood = round(sum(`Overall
effectiveness`=="good", na.rm = TRUE)/N*100,1),
propRI = round(sum(`Overall
effectiveness`=="requiresImprovement", na.rm = TRUE)/N*100,1),
propInadequate = round(sum(`Overall
effectiveness`=="inadequate", na.rm = TRUE)/N*100,1),

natMeanKS1 = round(mean(TKS1AVERAGE, na.rm =
TRUE),1),
natMeanFSM = round(mean(PTFSM6CLA1A, na.rm =
TRUE),1),
natMeanNonFSM = round(mean(100-PTFSM6CLA1A, na.rm =
TRUE),1),
natMeanEAL = round(mean(PTEALGRP2, na.rm =

```

```

natMeanNonEAL = round(mean(100-PTEALGRP2, na.rm =
TRUE),1))]

#-----PUPIL-LEVEL CHARACTERISTICS-----

# AS RANDOMISED

tablePupilChars_randomisation <- as.data.frame(t(data[, .(
  FSM = sprintf('%2.2f , %d/%d (%d) , %d (%2.1f)',
    nationalFigures$natMeanFSM, .N, data[, .N], sum(is.na(fsm)),
sum(fsm==1), round(mean(fsm==1)*100, 1)),
  nonFSM = sprintf('%2.2f , %d/%d (%d) , %d (%2.1f)',
    100-nationalFigures$natMeanFSM, .N, data[, .N], sum(is.na(fsm)),
sum(fsm==0), round(mean(fsm==0)*100, 1)),
  EAL = sprintf('%2.2f , %d/%d (%d) , %d (%2.1f)',
    nationalFigures$natMeanEAL, .N, data[, .N], sum(is.na(eal)),
sum(eal==1), round(mean(eal==1)*100, 1)),
  nonEAL = sprintf('%2.2f , %d/%d (%d) , %d (%2.1f)',
    100-nationalFigures$natMeanEAL, .N, data[, .N], sum(is.na(eal)),
sum(eal==0), round(mean(eal==0)*100, 1)),
  year1 = sprintf(' ---- , %d/%d (%d) , %d (%2.1f)',
    .N, data[, .N], sum(is.na(year)), sum(year==0),
round(mean(year==0)*100, 1)),
  year2 = sprintf(' ---- , %d/%d (%d) , %d (%2.1f)',
    .N, data[, .N], sum(is.na(year)), sum(year==1),
round(mean(year==1)*100, 1))),
  by = treat[order(-treat)])

# AS ANALYSED

tablePupilChars_analysis <- sapply(c("piraScoreNorm","bsScore"), function(outcome) {
  as.data.frame(t(data[, .(
    FSM = sprintf('%2.2f , %d/%d (%d) , %d (%2.1f)',
      nationalFigures$natMeanFSM, sum(!is.na(get(outcome))),
data[!is.na(get(outcome)), .N], sum(is.na(get(outcome))), sum(!is.na(get(outcome)) &
fsm==1), round(sum(!is.na(get(outcome)) & fsm==1)/sum(!is.na(get(outcome)))*100, 1)),
    nonFSM = sprintf('%2.2f , %d/%d (%d) , %d (%2.1f)',
      100-nationalFigures$natMeanFSM, sum(!is.na(get(outcome))),
data[!is.na(get(outcome)), .N], sum(is.na(get(outcome))), sum(!is.na(get(outcome)) &
fsm==0), round(sum(!is.na(get(outcome)) & fsm==0)/sum(!is.na(get(outcome)))*100, 1)),
    EAL = sprintf('%2.2f , %d/%d (%d) , %d (%2.1f)',
      nationalFigures$natMeanEAL, sum(!is.na(get(outcome))),
data[!is.na(get(outcome)), .N], sum(is.na(get(outcome))), sum(!is.na(get(outcome)) &
eal==1), round(sum(!is.na(get(outcome)) & eal==1)/sum(!is.na(get(outcome)))*100, 1)),
    nonEAL = sprintf('%2.2f , %d/%d (%d) , %d (%2.1f)',
      100-nationalFigures$natMeanEAL, sum(!is.na(get(outcome))),
data[!is.na(get(outcome)), .N], sum(is.na(get(outcome))), sum(!is.na(get(outcome)) &
eal==0), round(sum(!is.na(get(outcome)) & eal==0)/sum(!is.na(get(outcome)))*100, 1)),
    year1 = sprintf(' ---- , %d/%d (%d) , %d (%2.1f)',
      sum(!is.na(get(outcome))), data[!is.na(get(outcome)), .N],
sum(is.na(get(outcome))), sum(!is.na(get(outcome)) & year==0),
round(sum(!is.na(get(outcome)) & year==0)/sum(!is.na(get(outcome)))*100, 1)),
    year2 = sprintf(' ---- , %d/%d (%d) , %d (%2.1f)',
      sum(!is.na(get(outcome))), data[!is.na(get(outcome)), .N],
sum(is.na(get(outcome))), sum(!is.na(get(outcome)) & year==1),
round(sum(!is.na(get(outcome)) & year==1)/sum(!is.na(get(outcome)))*100, 1))),
  by = treat[order(-treat)])
}, simplify = FALSE)

#-----SCHOOL-LEVEL CHARACTERISTICS-----

```

```
# trial school data
dataSchool <- as.data.table(read.dta13(PATH_TO_SCHOOLS_LIST))

dataSchool <- dataSchool[!duplicated(dataSchool)]
dataSchool[, URN := as.character(urn)]

# merge with national data
dataSchool <- Reduce(function(x,y) merge(x, y, by = "URN", all.x = TRUE),
list(dataSchool,dataSpine,dataKS2,dataOfsted,dataEdubase))

dataSchool[, schooltype := case_when(NFTYPE %in% c("Academy - Converter mainstream",
"Academy sponsor led mainstream",
"Free school - Mainstream") ~
"academy",
NFTYPE=="Community school" ~ "community",
NFTYPE %in% c("Foundation school",
"Voluntary aided school",
"Voluntary controlled school") ~
"other")]

# school-level characteristics - as randomised and as analysed (excluding 1 school)
tableSchoolChars <- sapply(c("randomised" = "null", "analysed" = "Alt Academy"),
function(schoolToExclude){
  t(dataSchool[school != schoolToExclude,
  .(urban = sprintf('%2.2f , %2.0f (%2.0f) , %2.0f (%2.1f)',
nationalFigures$propUrban, .N,
sum(UrbanRural..name==""),
sum(grepl("Urban",UrbanRural..name., ignore.case =
TRUE))),
round(sum(grepl("Urban",UrbanRural..name., ignore.case
= TRUE))/N*100,1)),
rural = sprintf('%2.2f , %2.0f (%2.0f) , %2.0f (%2.1f)',
nationalFigures$propRural, .N,
sum(UrbanRural..name==""),
sum(!grepl("Urban",UrbanRural..name., ignore.case =
TRUE))),
round(sum(!grepl("Urban",UrbanRural..name.,
ignore.case = TRUE))/N*100,1)),
community = sprintf('%2.2f , %2.0f (%2.0f) , %2.0f (%2.1f)',
nationalFigures$propCommunity, .N,
sum(is.na(schooltype)),
sum(schooltype=="community"),
round(sum(schooltype=="community")/N*100,1)),
academy = sprintf('%2.2f , %2.0f (%2.0f) , %2.0f (%2.1f)',
nationalFigures$propAcademy, .N,
sum(is.na(schooltype)),
sum(schooltype=="academy"),
round(sum(schooltype=="academy")/N*100,1)),
other = sprintf('%2.2f , %2.0f (%2.0f) , %2.0f
(%2.1f)', nationalFigures$propOther, .N,
sum(is.na(schooltype)),
sum(schooltype=="other"),
round(sum(schooltype=="other")/N*100,1)),
outstanding = sprintf('%2.2f , %2.0f (%2.0f) , %2.0f (%2.1f)',
nationalFigures$propOutstanding, .N,
sum(is.na(`Overall effectiveness`)),
sum(`Overall effectiveness`=="outstanding",
na.rm = TRUE),
round(sum(`Overall
effectiveness`=="outstanding", na.rm = TRUE)/N*100,1)),
good = sprintf('%2.2f , %2.0f (%2.0f) , %2.0f (%2.1f)',
nationalFigures$propGood, .N,
```

```

      sum(is.na(`Overall effectiveness`)),
      sum(`Overall effectiveness`=="good", na.rm = TRUE),
      round(sum(`Overall effectiveness`=="good", na.rm =
TRUE)/.N*100,1)),
      reqImp = sprintf('%2.2f , %2.0f (%2.0f) , %2.0f (%2.1f)',
nationalFigures$propRI, .N,
      sum(is.na(`Overall effectiveness`)),
      sum(`Overall effectiveness`=="requiresImprovement",
na.rm = TRUE),
      round(sum(`Overall
effectiveness`=="requiresImprovement", na.rm = TRUE)/.N*100,1)),
      ks1 = sprintf('%2.2f , %2.0f (%2.0f) , %2.1f (%2.1f)',
nationalFigures$natMeanKS1, .N,
      sum(is.na(TKS1AVERAGE)),
      round(mean(TKS1AVERAGE, na.rm = TRUE),1),
      round(sd(TKS1AVERAGE, na.rm = TRUE),1)),
      fsm = sprintf('%2.2f , %2.0f (%2.0f) , %2.1f (%2.1f)',
nationalFigures$natMeanFSM, .N,
      sum(is.na(PTFSM6CLA1A)),
      round(mean(PTFSM6CLA1A, na.rm = TRUE),1),
      round(sd(PTFSM6CLA1A, na.rm = TRUE),1)),
      eal = sprintf('%2.2f , %2.0f (%2.0f) , %2.1f (%2.1f)',
nationalFigures$natMeanEAL, .N,
      sum(is.na(PTEALGRP2)),
      round(mean(PTEALGRP2, na.rm = TRUE),1),
      round(sd(PTEALGRP2, na.rm = TRUE),1)))]]
}, simplify = FALSE)

```

```
#-----OVERALL RAW MEANS AND SDs-----
```

```

# raw means + SD
overallMeanSD <- dataLong[!is.na(outcomeValue),
  .(mean = round(mean(outcomeValue),2),
    sd = round(sd(outcomeValue),2)),
  by = .(outcomeType)][order(outcomeType)]

```

```
#-----PLOT PRIMARY AND SECONDARY OUTCOME HISTOGRAMS-----
```

```

binwidth <- c(0.4,4,1,4)
bincentre <- c(0.2,2,1,2)
xlims <- list(c(-3,4),c(0,60),c(2,10),c(0,140))
ylims <- list(c(0,150),c(0,150),c(0,300),c(0,120))
xbreaks <- list(seq(-3,4,1),seq(0,60,4),seq(3,9,1),seq(0,138,10))
ybreaks <- list(seq(0,150,25),seq(0,150,25),seq(0,300,50),seq(0,120,20))
xtitle <- list("PIRA score (normalised per year group)", "Renfrew Bus Story
score", "WSEM3 score", "SSIS score")

```

```

for (plotX in c("primary","secondary")) {
  outcomes <- case_when(plotX=="primary" ~ c("piraScoreNorm","bsScore"),
    plotX=="secondary" ~ c("wsem3score","ssisScore"))
  plotOutput <- sapply(outcomes, function(outcomeX){
    idx <- match(outcomeX,c("piraScoreNorm","bsScore","wsem3score","ssisScore"))
    plotOutput <- ggplot() +
      # histogram
      geom_histogram(data = data[!is.na(get(outcomeX)),],

```

```

        aes(x = get(outcomeX)), binwidth = binwidth[idx], center =
bincentre[idx], color = "white", fill = "gray") +
        scale_x_continuous(limits = xlims[[idx]], breaks = xbreaks[[idx]], name =
xtitle[[idx]], expand = c(0.01, 0)) +
        scale_y_continuous(limits = ylims[[idx]], breaks = ybreaks[[idx]], name =
"Count", expand = c(0.01, 0)) +

        theme(panel.background = element_blank(),
              axis.line = element_line(),
              axis.text = element_text(size = 12),
              axis.title = element_text(size = 12, face = "bold"))

    return(plotOutput)
  }, simplify = FALSE)

# arrange plots
plotPrint <- ggarrange(plotOutput[[1]],plotOutput[[2]],
                      ncol = 2, nrow = 1, common.legend = FALSE, align = "hv")

# save
ggsave(paste(outputDir, sprintf("plot_histogram_%s_outcomes.png",plotX), sep="/"), plot
= plotPrint, device = NULL, path = NULL,
      dpi = 300, width = 28, height = 10, units = "cm", limitsize = TRUE)
}

#-----FUNCTION TO COMPUTE HEDGE'S G-----
hedges_g <- function(treatEffect, dataTreat, dataControl){

  # get Ns per arm
  Ncontrol <- sum(!is.na(dataControl))
  Ntreat <- sum(!is.na(dataTreat))

  # compute SD per arm
  SDcontrol <- sd(dataControl, na.rm = TRUE)
  SDtreat <- sd(dataTreat, na.rm = TRUE)

  # compute pooled SD
  SDpooled <- sqrt(((Ncontrol-1)*(SDcontrol^2) + (Ntreat-1)*(SDtreat^2)) /
(Ncontrol+Ntreat-2))

  # compute correction factor using gamma dist; if intractable, use approximation as
per SAP
  corrFactor <- gamma((Ncontrol+Ntreat+2)/2) / (sqrt((Ncontrol+Ntreat+2)/2) *
gamma((Ncontrol+Ntreat+2-1)/2))
  corrFactor <- ifelse(is.nan(corrFactor), 1 - (3/(4*(Ncontrol+Ntreat)-9)), corrFactor)

  return(list(SDcontrol = SDcontrol,
            SDtreat = SDtreat,
            g = corrFactor * (treatEffect / SDpooled),
            corrFactor = corrFactor,
            SDpooled = SDpooled))
}

#-----PRIMARY AND SECONDARY ANALYSIS-----

# run primary and secondary analysis
out <- sapply(c("piraScoreNorm", "bsScore", "wsem3score", "ssisScore"), function(outcome)
{

  # run GLM

```

```

glmOut <- glm(formula(sprintf('%s ~ treat + fsm + eal + year +
anonschoolid', outcome)),
              data = data[!is.na(get(outcome)),], family = "gaussian")

# get treatment effect
treatEffect <- glmOut$coefficients["treat1"]

# compute robust SE (not clustered)
glmOut_robust <- as.data.table(tidy(lmtest::coefstest(glmOut, vcov =
sandwich::vcovHC(glmOut))))

# compute robust CI (Bonferroni corrected to 97.5% if primary outcomes, 95%
otherwise)
CI <- qnorm(ifelse(outcome %in% c("piraScoreNorm", "bsScore"), 0.9875, 0.975)) *
glmOut_robust[term=="treat1", std.error]

# compute cluster-robust SE and CI using wild bootstrapping
CI_level <- ifelse(outcome %in% c("piraScoreNorm", "bsScore"), 0.975, 0.95)
glmOut_clus <- cluster.wild.glm(glmOut, cluster = ~ anonschoolid,
                               dat = data[!is.na(get(outcome)),],
                               boot.reps = nbootreps,
                               impose.null = FALSE,
                               ci.level = CI_level,
                               report = FALSE)

# compute hedge's g and CI
gIn <- c(treatEffect, treatEffect-CI, treatEffect+CI,
         glmOut_clus$ci["treat1", "CI lower"],
         glmOut_clus$ci["treat1", "CI higher"])

gOut <- sapply(gIn, function(x) {
  hedges_g(treatEffect = x,
           dataTreat = data[treat==1 & !is.na(get(outcome)), get(outcome)],
           dataControl = data[treat==0 & !is.na(get(outcome)), get(outcome)])
}, simplify = FALSE)

return(list(# main table output
  N_missing_treat = sprintf('%d (%d)', data[treat==1 & !is.na(get(outcome)), .N],
data[treat==1 & is.na(get(outcome)), .N]),
  unadjMeanCI_Treat = data[!is.na(get(outcome)) & treat==1,
                           sprintf('%2.2f (%2.2f, %2.2f)', mean(get(outcome)),
mean(get(outcome)) -
qnorm(0.975)*(sd(get(outcome))/sqrt(.N)),
mean(get(outcome)) +
qnorm(0.975)*(sd(get(outcome))/sqrt(.N)))],
  N_missing_control = sprintf('%d (%d)', data[treat==0 & !is.na(get(outcome)), .N],
data[treat==0 & is.na(get(outcome)), .N]),
  unadjMeanCI_Control = data[!is.na(get(outcome)) & treat==0,
                              sprintf('%2.2f (%2.2f, %2.2f)', mean(get(outcome)),
mean(get(outcome)) -
qnorm(0.975)*(sd(get(outcome))/sqrt(.N)),
mean(get(outcome)) +
qnorm(0.975)*(sd(get(outcome))/sqrt(.N)))],
  total_N = sprintf('%d (%d, %d)', data[!is.na(get(outcome)), .N],
                    data[treat==1 & !is.na(get(outcome)), .N],
                    data[treat==0 & !is.na(get(outcome)), .N]),
  hedges_g_CI = sprintf('%2.2f (%2.2f, %2.2f)', gOut[[1]]$g, gOut[[2]]$g,
gOut[[3]]$g),
  p_value = round(ifelse(outcome %in% c("piraScoreNorm", "bsScore"),
                        min(2*glmOut_robust[term=="treat1", p.value], 1),

```

```

      glmOut_robust[term=="treat1", p.value]),2), # Bonferroni
correct the 2 primary outcomes

# effect size table output
unadjDiff = round(data[!is.na(get(outcome)) & treat==1, mean(get(outcome))] -
                  data[!is.na(get(outcome)) & treat==0, mean(get(outcome))],2),
adjDiff = round(unnamed(treatEffect),2),
sdTreat = round(gOut[[1]]$SDtreat,2),
sdControl = round(gOut[[1]]$SDcontrol,2),
SDpooled = round(gOut[[1]]$SDpooled,2),
corrFactor = round(gOut[[1]]$corrFactor,5),

# clustered SE robustness check
hedges_g_CI_clus = sprintf('%2.2f (%2.2f, %2.2f)', gOut[[1]]$g, gOut[[4]]$g,
gOut[[5]]$g),
p_clus = round(iffelse(outcome %in% c("piraScoreNorm","bsScore"),
                      min(2*glmOut_clus$p.values["treat1","wild cluster BS p-
value"],1),
                      glmOut_clus$p.values["treat1","wild cluster BS p-
value"]),2)))

}, simplify = TRUE)

# primary/secondary results tables
tablePrimaryResults <- data.frame(t(out[1:7,1:2]))
tableSecondaryResults <- data.frame(t(out[1:7,3:4]))

# effect size table (annex)
tableEffectSizes <- data.frame(t(out[c(8,9,1,10,3,11:13),]))

# clustered SE robustness check table
tableClusteredSEcheck <- data.frame(t(out[c(5,14:15),]))

#-----COMPLIANCE DATA SUMMARY-----

# proportion compliers (excl missing compliance from denominator)
# 1 control pupil attended 22 sessions (hence non-zero mean)
complianceSummary <- data[, .(propAvailCompliance =
round(mean(!is.na(isComplier))*100,1),
                        N_availCompliance = sum(!is.na(isComplier)),
                        propMissingCompliance =
round(mean(is.na(isComplier))*100,1),
                        N_missingCompliance = sum(is.na(isComplier)),
                        propCompliers_exclMissing = round(mean(isComplier, na.rm
= TRUE)*100,1),
                        meanAttended = mean(sessAttended, na.rm = TRUE),
                        medianAttended = median(sessAttended, na.rm = TRUE)), by
= treat]

# histogram of sessions attended
plotOut <- ggplot() +

# histogram
geom_histogram(data = data[!is.na(sessAttended) & treat==1,],
              aes(x = sessAttended), binwidth = 1, center = 1, color = "white", fill
= "gray") +
scale_x_continuous(limits = c(0,25), breaks = seq(0,24,2),
                  name = 'Number of Speech Bubbles sessions attended', expand =
c(0.01, 0)) +
scale_y_continuous(limits = c(0,120), breaks = seq(0,120,20), name = "Count", expand
= c(0.01, 0)) +

```

```

# indicate threshold for binary compliance
geom_vline(xintercept = 16, color = "black", linetype = "dashed") +
  annotate("text", x = 16.3, y = 110, label = "minimum number of \nsessions for
compliance",
  size = 3.5, hjust = 0, color = "black") +

theme(panel.background = element_blank(),
  axis.line = element_line(),
  axis.text = element_text(size = 12),
  axis.title = element_text(size = 12, face = "bold"))

# save
ggsave(paste(outputDir,"plot_histogram_sessAttended.png",sep="/"), plot = plotOut,
device = NULL, path = NULL,
  dpi = 300, width = 14, height = 10, units = "cm", limitsize = TRUE)

#-----CACE ANALYSIS-----

# Ns for CACE analysis
data[!is.na(isComplier) & !is.na(piraScoreNorm), .N, by = treat]
data[!is.na(isComplier) & !is.na(bsScore), .N, by = treat]

# run CACE analysis for all outcomes
out <- sapply(c("piraScoreNorm","bsScore","wsem3score","ssisScore"), function(outcome)
{

  # first stage
  glmOut_firststage <- glm(isComplier ~ treat + fsm + eal + year + anonschoolid,
    data = data[!is.na(isComplier) & !is.na(get(outcome)),],
  family = "gaussian")

  # compute cluster-robust SE and CI using wild bootstrapping
  glmOut_firststage_robust <- as.data.table(tidy(lmtest::coefest(glmOut_firststage,
vcov = sandwich::vcovHC(glmOut_firststage))))

  # corr between compliance and treatment
  corrOut <- cor.test(data[!is.na(isComplier) &
!is.na(get(outcome)),as.numeric(isComplier)],
    data[!is.na(isComplier) &
!is.na(get(outcome)),as.numeric(treat)])

  # IV regression: regress compliance (instrument) onto treatment and then predicted
treatment onto outcome
  glmOut <- ivreg(get(outcome) ~ isComplier + fsm + eal + year + anonschoolid | treat +
fsm + eal + year + anonschoolid,
    data = data[!is.na(isComplier) & !is.na(get(outcome)),])

  # save diagnostics to get first stage partial F-test (below)
  ivregOut <- summary(glmOut, diagnostics = TRUE)

  # get treatment effect
  treatEffect <- glmOut$coefficients["isComplierTRUE"]

  # compute robust SE
  glmOut_robust <- as.data.table(tidy(lmtest::coefest(glmOut, vcov =
sandwich::vcovHC(glmOut))))

  # compute robust 95% CI
  CI <- qnorm(0.975) * glmOut_robust[term=="isComplierTRUE",std.error]

  # compute hedge's g and CI
  gIn <- c(treatEffect, treatEffect-CI, treatEffect+CI)

```

```

gOut <- sapply(gIn, function(x) {
  hedges_g(treatEffect = x,
    dataTreat = data[treat==1 & (!is.na(isComplier) | !is.na(get(outcome))),
get(outcome)],
    dataControl = data[treat==0 & (!is.na(isComplier) | !is.na(get(outcome))),
get(outcome)])
  }, simplify = FALSE)

  return(list(N_missing_treat = sprintf('%d (%d)',
    data[treat==1 & !is.na(get(outcome)) &
!is.na(isComplier), .N],
    data[treat==1 & (is.na(get(outcome)) |
is.na(isComplier)), .N]),
    N_missing_control = sprintf('%d (%d)',
    data[treat==0 & !is.na(get(outcome)) &
!is.na(isComplier), .N],
    data[treat==0 & (is.na(get(outcome)) |
is.na(isComplier)), .N]),
    weakInst_ftest = sprintf('F(%d, %d) = %2.2f (p = %2.3f)',
    ivregOut$diagnostics["Weak instruments", "df1"],
    ivregOut$diagnostics["Weak instruments", "df2"],
    ivregOut$diagnostics["Weak instruments",
"statistic"],
    ivregOut$diagnostics["Weak instruments", "p-
value"]),
    compTreat_corr = sprintf('%2.2f (p = %2.3f)', corrOut$estimate,
corrOut$p.value),
    hedges_g_CI = sprintf('%2.2f (%2.2f, %2.2f)', gOut[[1]]$g, gOut[[2]]$g,
gOut[[3]]$g),
    p = round(glmOut_robust[term=="isComplierTRUE", p.value], 2))

}, simplify = TRUE)

# primary/secondary CACE results tables
tableCACEResults <- data.frame(t(out))

#-----FSM INTERACTION ANALYSIS-----
out <- sapply(c("piraScoreNorm", "bsScore", "wsem3score", "ssisScore"), function(outcome)
{
  # run GLM
  glmOut <- glm(formula(sprintf('%s ~ fsm*treat + eal + year + anonschoolid', outcome)),
    data = data[!is.na(get(outcome)), ], family = "gaussian")

  # get interaction effect
  treatEffect <- glmOut$coefficients["fsm1:treat1"]

  # compute robust SE
  glmOut_robust <- as.data.table(tidy(lmtest::coefest(glmOut, vcov =
sandwich::vcovHC(glmOut))))

  # compute robust 95% CI
  CI <- qnorm(0.975) * glmOut_robust[term=="fsm1:treat1", std.error]

  # compute hedge's g and CI
  gIn <- c(treatEffect, treatEffect-CI, treatEffect+CI)

  gOut <- sapply(gIn, function(x) {
    hedges_g(treatEffect = x,
      dataTreat = data[treat==1 & !is.na(get(outcome)), get(outcome)],

```

```

        dataControl = data[treat==0 & !is.na(get(outcome)), get(outcome)]
    }, simplify = FALSE)

    return(list(hedges_g_CI = sprintf('%2.2f (%2.2f, %2.2f)', gOut[[1]]$g, gOut[[2]]$g,
gOut[[3]]$g),
        p_value = round(glmOut_robust[term=="fsm1:treat1", p.value],2)))

}, simplify = TRUE)

# results table
tableFSMinteraction <- data.frame(t(out))

#-----FSM SUBGROUP ANALYSIS-----

out <- sapply(c("piraScoreNorm","bsScore","wsem3score","ssisScore"), function(outcome)
{

    # run GLM (FSM only)
    glmOut <- glm(formula(sprintf('%s ~ treat + eal + year + anonschoolid',outcome)),
        data = data[fsm==1 & !is.na(get(outcome)),], family = "gaussian")

    # get interaction effect
    treatEffect <- glmOut$coefficients["treat1"]

    # compute robust SE
    glmOut_robust <- as.data.table(tidy(lmtest::coefTest(glmOut, vcov =
sandwich::vcovHC(glmOut))))

    # compute robust 95% CI
    CI <- qnorm(0.975) * glmOut_robust[term=="treat1",std.error]

    # compute hedge's g and CI
    gIn <- c(treatEffect, treatEffect-CI, treatEffect+CI)

    gOut <- sapply(gIn, function(x) {
        hedges_g(treatEffect = x,
            dataTreat = data[fsm==1 & treat==1 & !is.na(get(outcome)), get(outcome)],
            dataControl = data[fsm==1 & treat==0 & !is.na(get(outcome)),
get(outcome)])
    }, simplify = FALSE)

    return(list(total_N = sprintf('%d (%d, %d)',
        data[fsm==1 & !is.na(get(outcome)), .N],
        data[fsm==1 & treat==1 & !is.na(get(outcome)), .N],
        data[fsm==1 & treat==0 & !is.na(get(outcome)), .N]),
        hedges_g_CI = sprintf('%2.2f (%2.2f, %2.2f)', gOut[[1]]$g, gOut[[2]]$g,
gOut[[3]]$g),
        p_value = round(glmOut_robust[term=="treat1", p.value],2)))

}, simplify = TRUE)

# results table
tableFSMsubgroup <- data.frame(t(out))

#-----EAL INTERACTION ANALYSIS-----

out <- sapply(c("piraScoreNorm","bsScore","wsem3score","ssisScore"), function(outcome)
{

    # run GLM
    glmOut <- glm(formula(sprintf('%s ~ eal*treat + fsm + year + anonschoolid',outcome)),
        data = data[!is.na(get(outcome)),], family = "gaussian")

```

```

# get interaction effect
treatEffect <- glmOut$coefficients["eall:treat1"]

# compute robust SE
glmOut_robust <- as.data.table(tidy(lmtest::coefest(glmOut, vcov =
sandwich::vcovHC(glmOut))))

# compute robust 95% CI
CI <- qnorm(0.975) * glmOut_robust[term=="eall:treat1",std.error]

# compute hedge's g and CI
gIn <- c(treatEffect, treatEffect-CI, treatEffect+CI)

gOut <- sapply(gIn, function(x) {
  hedges_g(treatEffect = x,
    dataTreat = data[treat==1 & !is.na(get(outcome)), get(outcome)],
    dataControl = data[treat==0 & !is.na(get(outcome)), get(outcome)])
}, simplify = FALSE)

return(list(hedges_g_CI = sprintf('%2.2f (%2.2f, %2.2f)', gOut[[1]]$g, gOut[[2]]$g,
gOut[[3]]$g),
  p_value = round(glmOut_robust[term=="eall:treat1", p.value],2))
), simplify = TRUE)

# results table
tableEALinteraction <- data.frame(t(out))

#-----EAL SUBGROUP ANALYSIS-----

# run analysis for both EAL and non-EAL subgroups
out <- sapply(c("EAL" = 1, "Not EAL" = 0), function(isEAL) {

  sapply(c("piraScoreNorm", "bsScore", "wsem3score", "ssisScore"), function(outcome) {

    # check which schools have at least 1 EAL or 1 non-EAL pupil in both arms
    tmp <- data[eal==isEAL & !is.na(get(outcome)), .(bothArmsEAL = sum(treat==1) > 0 &
sum(treat==0) > 0), by = .(anonschoolid)]

    # run GLM (EAL/not EAL only)
    glmOut <- glm(formula(sprintf('%s ~ treat + fsm + year + anonschoolid',outcome)),
      data = data[eal==isEAL & !is.na(get(outcome)) & anonschoolid %in%
tmp[bothArmsEAL==TRUE, anonschoolid] ], family = "gaussian")

    # get treatment effect
    treatEffect <- glmOut$coefficients["treat1"]

    # compute robust SE
    glmOut_robust <- as.data.table(tidy(lmtest::coefest(glmOut, vcov =
sandwich::vcovHC(glmOut))))

    # compute robust 95% CI
    CI <- qnorm(0.975) * glmOut_robust[term=="treat1",std.error]

    # compute hedge's g and CI
    gIn <- c(treatEffect, treatEffect-CI, treatEffect+CI)

    gOut <- sapply(gIn, function(x) {
      hedges_g(treatEffect = x,
        dataTreat = data[eal==isEAL & treat==1

```

```

      & anonschoolid %in% tmp[bothArmsEAL==TRUE,
anonschoolid] & !is.na(get(outcome)), get(outcome)],
      dataControl = data[eal==isEAL & treat==0
      & anonschoolid %in% tmp[bothArmsEAL==TRUE,
anonschoolid] & !is.na(get(outcome)), get(outcome)]
    }, simplify = FALSE)

    return(list(total_N = sprintf('%d (%d, %d)',
      data[eal==isEAL & anonschoolid %in%
tmp[bothArmsEAL==TRUE, anonschoolid] & !is.na(get(outcome)), .N],
      data[eal==isEAL & treat==1 & anonschoolid %in%
tmp[bothArmsEAL==TRUE, anonschoolid] & !is.na(get(outcome)), .N],
      data[eal==isEAL & treat==0 & anonschoolid %in%
tmp[bothArmsEAL==TRUE, anonschoolid] & !is.na(get(outcome)), .N]),
      hedges_g_CI = sprintf('%2.2f (%2.2f, %2.2f)', gOut[[1]]$g, gOut[[2]]$g,
gOut[[3]]$g),
      p_value = round(glmOut_robust[term=="treat1", p.value],2))

  }, simplify = TRUE)
}, simplify = FALSE)

# results table
tableEALsubgroup <- data.frame(t(out$EAL))
tableNonEALsubgroup <- data.frame(t(out$`Not EAL`))

#-----RANDOMISATION INFERENCE ROBUSTNESS CHECK-----
-----

# make copy of dataset to edit (RI package prefers chars instead of factors)
tmpData <- data
tmpData[, anonschoolid := as.character(anonschoolid)]
tmpData[, fsm := as.character(fsm)]

# create randomisation block id from school and year group
tmpData[, rblock := ifelse(year==0,paste0(anonschoolid,"_year0"),
paste0(anonschoolid,"_year1"))]

# loop over outcomes
resultsRI <- sapply(c("piraScoreNorm","bsScore","wsem3score","ssisScore"),
function(outcome) {
  out <- conduct_ri(formula(sprintf('%s ~ treat + fsm + eal + year +
anonschoolid',outcome)),
    assignment = "treat",
    outcome = outcome,
    declaration = declare_ra(N = tmpData[!is.na(get(outcome)), .N],
      blocks = tmpData[!is.na(get(outcome)),
rblock]),
    data = tmpData[!is.na(get(outcome)),],
    sims = 100)
  return(list(estimate = tidy(out)$estimate, p = tidy(out)$p.value))
}, simplify = TRUE)

#-----MISSING DATA CHECK-----

# for primary outcomes, nothing (other than individual school fixed effects) predicts
missingness
(glmOut_predictMissingness <- sapply(c("piraScoreNorm","bsScore"), function(outcome) {
  tmpOut <- tidy(glm(is.na(get(outcome)) ~ treat + fsm + eal + year + anonschoolid,
data = data, family = "binomial"))
  tmpOut <- tmpOut[c(1:5,16,24:30,6:15,17:23), c("term","p.value")]
  tmpOut$`p.value` <- round(tmpOut$`p.value`,3)
  return(tmpOut)
})

```

```
}, simplify = FALSE))

#-----ICC-----

# one per outcome
(ICC <- sapply(c("piraScoreNorm", "bsScore", "wsem3score", "ssisScore"), function(outcome)
{
  # estimate variance components model with school-level random effect
  vc <- as.data.table(VarCorr(lmer(get(outcome) ~ (1|anonschoolid), data)))

  # compute ICC
  return(round(vc[grp=="anonschoolid",vcov] / vc[, sum(vcov)],2))
}))
```

## Appendix N: Interview guides

### N1. Teaching assistant interview guide

The interviews should last around 25 minutes. The timings given for each section are a guide - you may spend longer or shorter on each section. Lead questions are presented in bold, with potential follow-up questions presented in a non-bold typeface. As the interviews are semi-structured, not all questions need to be asked and they do not need to be asked in order. The interviewer should be responsive to what the interviewee, following the direction of the conversation and following-up with additional questions as needed.

Main objective	Purpose of section	Guide timings
1. Introductions	Explains the purpose and ground rules for the interview.	3 mins
2. Background context	Allows the participant an opportunity to settle into the interview, as well as providing some background to the school and the context in which the Speech Bubbles programme is being delivered.	5 mins
3. Delivery Experience	This section will focus on understanding the perceived quality of the intervention, as well as experiences of the programme's delivery, including barriers and facilitators to delivery.	5 mins
4. Pupil engagement and programme mechanisms	To explore the the teaching assistant's perception of pupils' engagement in the sessions and the positive and negative impact of the programme, together with the mechanisms that brought about any impact identified.	5 mins
5. Training and support	To understand teaching assistant's experience of the training they received to deliver the programme, and the support they have received to deliver the programme, including from the school's SLT and class teachers.	5 mins
6. Close	Thank you and close	2 mins
<b>Observation of SB session</b>		
7. Reflection following observation of SB session	Where possible, this will be an opportunity to follow-up on any areas of interest arising from the observation.	5 mins

1. Introductions	3 mins
<p>Introduction:</p> <ul style="list-style-type: none"> <li>● Introduce yourself</li> <li>● Introduce BIT and IOE – explain that we are independently evaluating the Speech Bubbles programme, which is one of five programmes that are part of the Cultural Learning programme that is jointly funded by the Education Endowment Foundation and Royal Society of Arts.</li> </ul> <p>Aims of this interview:</p> <p>We are here to learn more about how the Speech Bubbles programme has worked in your class. We're interested in what involvement you have had with the programme, what has helped the programme to work, and what the challenges have been. We'd also like to understand any impact the programme has had on your school, particularly pupils in the Speech Bubbles Session(s).</p> <p>This interview:</p> <ul style="list-style-type: none"> <li>● Should take no more than 25 minutes plus five minutes for reflection after the session.</li> <li>● Stress that you want to understand the world from your (the respondent's) point of view. No answers are right or wrong – and we are not here to judge the decisions made or views held by [the interviewee].</li> </ul> <p>Anonymity and privacy:</p> <ul style="list-style-type: none"> <li>● All information gathered will be in strict confidence, unless there are concerns about safeguarding, and no-one will be named in any subsequent write-up of this research.</li> <li>● Explain that if at any point they feel uncomfortable or prefer not to answer a specific question they can just say so.</li> <li>● Explain that it is their choice whether they take part in the interview and they can end the interview at any point, without giving a reason.</li> </ul> <p>Recording:</p> <ol style="list-style-type: none"> <li>1. Explain that recording enables us to have an accurate record of what was said, which can be typed up for analysis alongside other interviews. We may also use quotes from this interview, but these will be included in a way that means no individual or school is identifiable.</li> <li>2. Check if they have any questions about the interview. If they are happy to go ahead, obtain verbal permission to digitally record and take notes (written permission should already have been obtained).</li> <li>3. Once you have consent, start the voice recorder.</li> </ol>	<p>Orientates respondent and gets them prepared to take part in the discussion.</p> <p>Outlines the 'rules' of the interview.</p>

4. State interview number/-participant ID.	
<b>2. Background context</b>	<b>5 mins</b>
<p>Could you tell me a little bit about your role at the school?</p> <ul style="list-style-type: none"> <li>• What would you say are the school's main strengths and challenges?</li> </ul> <p>How did you become involved with the Speech Bubbles project?</p> <p>What were your thoughts about the Speech Bubbles programme when you first heard about it?</p> <ul style="list-style-type: none"> <li>• How did you feel about leading the speech bubbles project at your school?</li> <li>• What were you looking forward to about your taking part? [probe for individual feelings + feelings about student participation]</li> <li>• What challenges did you envisage?</li> </ul> <p>What experience of Drama do you have?</p>	<p>Allows the participant an opportunity to settle into the interview, as well as providing some background to the school and the context in which the Speech Bubbles programme was delivered.</p>
<b>3. Delivery Experience</b>	<b>5 mins</b>
<p>How have you found the Speech Bubbles sessions?</p> <ul style="list-style-type: none"> <li>• What has the Speech Bubbles programme consisted of?</li> <li>• What is the role and purpose of the Teaching Assistant within the Speech Bubbles programme?</li> <li>• What do you like about the Speech Bubbles session?</li> <li>• What have the main challenges been?</li> <li>• Anything that could have been done to help overcome these?</li> </ul> <p>How have you found making time to run the Speech Bubbles sessions every week?</p> <ul style="list-style-type: none"> <li>• When in the day do you typically run the sessions? Why is that?</li> <li>• How long do they typically last? What do you think about that length of session?</li> <li>• Where do the sessions typically take place? How have you found using this space for the sessions?</li> </ul> <p>How have you found working with the drama specialist?</p> <ul style="list-style-type: none"> <li>• How have they been with the pupils?</li> <li>• What was drama specialist's typical role within the Speech Bubbles sessions?</li> <li>• Did you face any challenges working with the drama specialist?</li> <li>• How did you overcome these challenges?</li> </ul>	<p>This section will focus on understanding the perceived quality of the intervention, as well as experiences of the programme's delivery, including barriers and facilitators to delivery.</p>

<ul style="list-style-type: none"> <li>To what extent have you been able to share your knowledge with one another?</li> </ul>	
<p><b>4. Pupil Engagement and Programme Mechanism</b></p>	<p><b>5 mins</b></p>
<p>How have children in your school found the Speech Bubbles sessions?</p> <ul style="list-style-type: none"> <li>Could you describe the pupils who have got the most out of the sessions? Why do you think this is?</li> <li>Could you describe any children who have struggled more with the sessions? Why do you think this is?</li> </ul> <p>What effects have the Speech Bubbles sessions had for your pupils?</p> <ul style="list-style-type: none"> <li>Can you describe any other changes you've noticed in your pupils?</li> <li>If not mentioned, probe for effect on classroom engagement and learning, and speech and communication. e.g. Can you describe any effect of the programme on pupils' communication skills?</li> <li>What is it about programme that you think has helped them to develop their (e.g. confidence)?</li> <li>Can you describe any negative consequences of the programme for pupils?</li> </ul> <p>Can you tell me a little more about how the students were referred to the Speech Bubbles programme?</p> <ul style="list-style-type: none"> <li>How appropriate were the student referrals?</li> <li>Knowing what you know now, are there any changes you would make to the pupil referral process?</li> </ul> <p>Is there anything you've learnt from taking part in the Speech Bubbles programme?</p>	<p>To explore the the teaching assistant's perception of pupils' engagement in the sessions and the positive and negative impact of the programme, together with the mechanisms that brought about any impact identified.</p>

<p>[probe for effect of participating in programme on TA e.g. professional development, confidence, role in the classroom, ownership, relationship with students]</p> <ul style="list-style-type: none"> <li>• What was it about the programme that you think led to these changes?</li> </ul>	
<p><b>5. Training and support</b></p>	<p><b>5 mins</b></p>
<p><i>If short for time, this section is lower priority than those above, training may have been covered in the previous sections and SLT/Teacher response is explored in SLT/Teacher interviews.</i></p> <p>Can you tell me about the training sessions that you attended as part of the Speech Bubbles programme?</p> <ul style="list-style-type: none"> <li>• What did you cover in the training sessions?</li> <li>• What worked well about the sessions?</li> <li>• What did you find challenging about the sessions?</li> <li>• To what extent did the sessions support you to deliver the Speech Bubbles programme?</li> <li>• If you could change anything about the training, what would it be?</li> </ul> <p>How have the SLT responded to the Speech Bubbles programme at [case study school]?</p> <ul style="list-style-type: none"> <li>• Can you describe any support you've received from SLT to deliver the programme?</li> <li>• Can you describe any resources they've made available to deliver the programme?</li> </ul> <p>How has the class teacher responded to the Speech Bubbles programme?</p>	<p>To understand teaching assistant's experience of the training they received to deliver the programme, and the support they have received to deliver the programme, including from the school's SLT and class teachers.</p>

<ul style="list-style-type: none"> <li>• Can you describe any support you've received from the class teacher to deliver the programme?</li> </ul> <p>How has the programme been perceived by parents of participating students?</p>	
<p><b>6. Close</b></p>	<p><b>2 mins</b></p>
<p>Overall, would you recommend the programme to other schools?</p> <ul style="list-style-type: none"> <li>• Why/Why not?</li> </ul> <p>If you had a magic wand, what one thing would you change about Speech Bubbles and why?</p> <p>Was there anything else that you were hoping to discuss that we haven't yet had a chance to talk about?</p> <p>Thank the interviewee for their time and reassure them of the confidentiality of their responses, as explained at the beginning of the interview.</p>	<p>Thank you and close</p>

<p><b>7. Reflections on observations</b></p>	<p><b>5 mins</b></p>
----------------------------------------------	----------------------

<p><i>Below are some exemplar questions, however, the interviewer may want to add some additional questions based on their observations. Probe for their perception of student engagement; implementer support; and mechanisms.</i></p> <p>How did you find the session I have just observed?</p> <ul style="list-style-type: none"> <li>● What did you understand your role to be within the session?</li> <li>● What do you think went well?</li> <li>● What do you think were the challenges?</li> </ul> <p><i>[Use the student evaluation sheet as a prompt]</i></p> <ul style="list-style-type: none"> <li>● Which students engaged particularly well? How/Why?</li> <li>● Which students engaged less well? How/Why?</li> <li>● How did that session compare to previous sessions?</li> <li>● Could you describe any ways that the sessions have changed over time?</li> <li>● What do you hope to do in future Speech Bubbles Sessions?</li> <li>● If you could go back and do that session again, what would do differently?</li> </ul>	<p>Opportunity to follow-up on any areas of interest arising from the observation.</p>
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------

## N2. Teacher interview guide

The interviews should last around 25 minutes. The timings given for each section are a guide - you may spend longer or shorter on each section. Lead questions are presented in bold, with potential follow-up questions presented in a non-bold typeface. As the interviews are semi-structured, not all questions need to be asked and they do not need to be asked in order. The interviewer should be responsive to what the interviewee, following the direction of the conversation and following-up with additional questions as needed.

Main objective	Purpose of section	Guide timings
1. Introductions	Explains the purpose and ground rules for the interview.	3 mins
2. Background context	Allows the participant an opportunity to settle into the interview, as well as providing some background to the school and the context in which the Speech Bubbles programme is being delivered.	3 mins

3. Engagement	To understand the extent to which the school, and particularly the class teacher, has engaged with and supported the programme.	5 mins
4. Delivery Experience	This section will focus on understanding the perceived quality of the intervention, as well as the experience of the programme's delivery, including barriers and facilitators to delivery. We will also try and disentangle what is 'business as normal' within the school in terms of communication skills support and drama, and whether this has changed as a result of the Speech Bubbles programme.	6 mins
5. Mechanisms	To explore the teacher's perception of the positive and negative impact of the programme, particularly for pupils, and the mechanisms that brought about any change identified.	6 mins
6. Close	Thank you and close.	2 mins

### Topic guide

<b>1. Introductions</b>	<b>3 mins</b>
<p>Introduction:</p> <ul style="list-style-type: none"> <li>● Introduce yourself</li> <li>● Introduce BIT and IOE– explain that we are independently evaluating the Speech Bubbles programme, which is one of five programmes that are part of the Cultural Learning programme that is jointly funded by the Education Endowment Foundation and Royal Society of Arts.</li> </ul> <p>Aims of this interview:</p> <p>We are here to learn more about how the Speech Bubbles programme has worked in your class. We're interested in what involvement you have had with the programme and what has helped the programme to work, and what the challenges have been. We'd also like to understand any impact the programme has had on your school, particularly pupils in your class.</p> <p>This interview:</p>	<p>Orientates respondent and gets them prepared to take part in the discussion.</p> <p>Outlines the 'rules' of the interview.</p>

<ul style="list-style-type: none"> <li>● Should take no more than 25 minutes</li> <li>● Stress that you want to understand the world from your (the respondent's) point of view. No answers are right or wrong – and we are not here to judge the decisions made or views held by the interviewee.</li> </ul> <p>Anonymity and privacy:</p> <ul style="list-style-type: none"> <li>● All information gathered will be in strict confidence, unless there are concerns about safeguarding, and no-one will be named in any subsequent write-up of this research.</li> <li>● Explain that if at any point they feel uncomfortable or prefer not to answer a specific question they can just say so.</li> <li>● Explain that it is their choice whether they take part in the interview and they can end the interview at any point, without giving a reason.</li> </ul> <p>Recording:</p> <ul style="list-style-type: none"> <li>● Explain that recording enables us to have an accurate record of what was said, which can be typed up for analysis alongside other interviews. We may also use quotes from this interview, but these will be included in a way that means no individual or school is identifiable.</li> <li>● Check if they have any questions about the interview. If they are happy to go ahead, obtain verbal permission to digitally record and take notes (written permission should already have been obtained).</li> <li>● Once you have consent, start the voice recorder.</li> <li>● State interview number/participant ID</li> </ul>	
<p><b>2. Background context</b></p>	<p><b>3 mins</b></p>
<p>How long have you been teaching?</p> <p>Could you tell me about your role at the school?</p> <ul style="list-style-type: none"> <li>● What brought you to the school?</li> <li>● How long have you been at the school?</li> </ul> <p>Could you tell me a little more about what the school is like?</p> <ul style="list-style-type: none"> <li>● What would you say are the schools' main strengths?</li> <li>● What would you say are some of the school's biggest challenges?</li> <li>● What do you like most about the school?</li> </ul>	<p>Allows the participant an opportunity to settle into the interview, as well as providing some background to the school and the context in which the Speech Bubbles programme is being delivered.</p>

3. Engagement	5 mins
<p>Can you tell me about your involvement with the Speech Bubbles programme?</p> <ul style="list-style-type: none"> <li>● Can you describe any support that you, or other members of SLT, have provided to the programme?</li> </ul> <p>Have you sat in on any of the Speech Bubbles sessions?</p> <p>If yes:</p> <ul style="list-style-type: none"> <li>● Can you describe any sessions you've sat in on?</li> <li>● What did you think of the session[s]?</li> </ul> <p>[probe for further exploration of strength and suggested areas for improvement]</p> <p>If no:</p> <ul style="list-style-type: none"> <li>● What do you understand about what the students have been doing in the Speech Bubbles sessions?</li> </ul>	<p>To understand the extent to which the school, and particularly the class teacher, has engaged with and supported the programme.</p>
4. Delivery Experience	6 mins
<p>Can you tell me a little bit about the support that is normally offered for students with communication and social skills difficulties at your school?</p> <ul style="list-style-type: none"> <li>● Have you continued to offer these activities alongside the Speech Bubbles programme?</li> <li>● What did you hope the Speech Bubbles programme would add?</li> </ul> <p>Could you tell me a little more about the school's pre-existing Drama provision?</p> <ul style="list-style-type: none"> <li>● What kind of drama activities would pupils in [intervention year] typically receive?</li> <li>● How do you see the Speech Bubbles programme fitting in with other drama activities in the school? [probe for similarity/difference]</li> </ul> <p>Thinking specifically about the delivery of the programme, how have you found having the Speech Bubbles programme running in your school?</p> <ul style="list-style-type: none"> <li>● What has worked well about the delivery of the programme?</li> <li>● What have been the main challenges to delivering the programme? [e.g. working with drama practitioner/external agencies/staff buy-in, facilities, resources]</li> <li>● What have you done to try and overcome these challenges?</li> <li>● How have you found working with the Drama Practitioner?</li> </ul>	<p>This section will focus on understanding the perceived quality of the intervention, as well as the experience of the programme's delivery, including barriers and facilitators to delivery. We will also try and disentangle what is 'business as normal' within the school in terms of communication skills support and drama, and whether this has changed as a result of the Speech Bubbles programme.</p>

- How has the school found accommodating the weekly Speech Bubbles session into the school timetable?
- Knowing what you know now, would you approach anything about the programme implementation differently?

How have the SLT responded to the Speech Bubbles programme?

- Can you describe any support you've received from SLT to deliver the programme?
- Can you describe any resources they've made available to deliver the programme?

5. Mechanisms	6 mins
<p>How did you decide which students to refer to the Speech Bubbles programme?</p> <ul style="list-style-type: none"> <li>Knowing what you know now, would you change anything about the referral process in the future?</li> </ul> <p>How have children in your class found the Speech Bubbles sessions?</p> <ul style="list-style-type: none"> <li>Could you describe the pupils who have got the most out of the sessions? Why do you think this is?</li> <li>Could you describe any children who have struggled more with the sessions? Why do you think this is?</li> </ul> <p>What effects have the Speech Bubbles sessions had for your pupils?</p> <p>[probe for effect on classroom engagement and learning, and speech and communication. e.g.Can you describe any effect of the programme on pupils' communication skills?]</p> <ul style="list-style-type: none"> <li>What was it about the programme that you think led to the effects you have described?</li> <li>Are you aware of any students that the programme has been particularly helpful/unhelpful for?</li> </ul> <p>What effect, if any, do you think the Speech Bubbles has had for the Teaching Assistant participating in Speech Bubbles?</p> <ul style="list-style-type: none"> <li>When did you notice these effects? [probe for professional development, confidence, role in classroom, ownership, relationship with teachers]</li> <li>What do you think it is about the Speech Bubbles programme that has led to these changes?</li> </ul> <p>Can you describe any effect of the programme more broadly within the school?</p> <ul style="list-style-type: none"> <li>Can you describe any effect of the programme on other pupils within the school?</li> <li>Can you describe any effect of the programme on other teachers within the school?</li> <li>Has [intervention class] doing Speech Bubbles affected drama/speech and language provision within the school more generally? If yes, how?</li> </ul> <p>How has the programme been perceived by parents of participating students?</p>	<p>To explore the teacher's perception of the positive and negative impact of the programme, particularly for pupils, and the mechanisms that brought about any change identified.</p>

6. Close	2 mins
<p>Overall, would you recommend the programme to other schools?</p> <ul style="list-style-type: none"> <li>Why/Why not?</li> </ul> <p>If you had a magic wand, what one thing would you change about Speech Bubbles and why?</p> <p>Was there anything else that you were hoping to discuss that we haven't yet had a chance to talk about?</p> <p>Thank the interviewee for their time and reassure them of the confidentiality of their responses, as explained at the beginning of the interview.</p>	Thank you and close

### N3. Senior leadership team interview guide

The interviews should last around 20 minutes. The timings given for each section are a guide - you may spend longer or shorter on each section. Lead questions are presented in bold, with potential follow-up questions presented in a non-bold typeface. As the interviews are semi-structured, not all questions need to be asked and they do not need to be asked in order. The interviewer should be responsive to what the interviewee, following the direction of the conversation and following-up with additional questions as needed.

Main objective	Purpose of section	Guide timings
1. Introductions	Explains the purpose and ground rules for the interview.	3 mins
2. Background context	Allows the participant an opportunity to settle into the interview, as well as providing some background to the school, so that we understand more about the context in which the Speech Bubbles programme is being delivered.	2 mins
2. Engagement	To understand the extent to which the school, and particularly the SLT, has engaged with and supported the programme.	4 mins
3. Delivery Experience	This section will focus on understanding the perceived quality of the intervention, as well as the experience of the programme's delivery, including barriers and facilitators to delivery. We will also try and disentangle what is 'business as normal' within the school in terms of	5 mins

	communication skills support and drama, and whether this has changed as a result of the Speech Bubbles programme.	
4. Mechanisms	To explore the SLT member's perception of the positive and negative impact of the programme, particularly for pupils, and the mechanisms that brought about any change identified.	4mins
6. Close	Thank you and close	2 mins

<b>1. Introductions</b>	<b>3 mins</b>
<p>Introduction:</p> <ul style="list-style-type: none"> <li>● Introduce yourself</li> <li>● Introduce BIT and IOE– explain that we are independently evaluating the Speech Bubbles programme, which is one of five programmes that are part of the Cultural Learning programme that is jointly funded by the Education Endowment Foundation and Royal Society of Arts.</li> </ul> <p>Aims of this interview:</p> <p>We are here to learn more about how the Speech Bubbles programme has worked in your school. We're interested in what involvement you have had with the programme and what has helped the programme to work, and what the challenges have been. We'd also like to understand any impact the programme has had on your school, particularly pupils in the Speech Bubbles session(s).</p> <p>This interview:</p> <ul style="list-style-type: none"> <li>● Should take no more than 20 minutes.</li> <li>● Stress that you want to understand the world from your (the respondent's) point of view. No answers are right or wrong – and we are not here to judge the decisions made or views held by the interviewee.</li> </ul> <p>Anonymity and privacy:</p> <ul style="list-style-type: none"> <li>● All information gathered will be in strict confidence, unless there are concerns about safeguarding, and no-one will be named in any subsequent write-up of this research.</li> </ul>	<p>Orientates respondent and gets them prepared to take part in the discussion.</p> <p>Outlines the 'rules' of the interview.</p>

<ul style="list-style-type: none"> <li>● Explain that if at any point they feel uncomfortable or prefer not to answer a specific question they can just say so.</li> <li>● Explain that it is their choice whether they take part in the interview and they can end the interview at any point, without giving a reason.</li> </ul> <p>Recording:</p> <ul style="list-style-type: none"> <li>● Explain that recording enables us to have an accurate record of what was said, which can be typed up for analysis alongside other interviews. We may also use quotes from this interview, but these will be included in a way that means no individual or school is identifiable.</li> <li>● Check if they have any questions about the interview. If they are happy to go ahead, obtain verbal permission to digitally record and take notes (written permission should already have been obtained).</li> <li>● Once you have consent, start the voice recorder.</li> <li>● State interview number/participant ID</li> </ul>	
<p><b>2. Background Context</b></p>	<p><b>2 mins</b></p>
<p>How long have you been working at the school?</p> <p>Can you tell me about your role at the school?</p> <ul style="list-style-type: none"> <li>● What brought you to the school?</li> <li>● How long you've been at the school?</li> </ul> <p>Could you tell me a little about the school itself?</p> <ul style="list-style-type: none"> <li>● What would you say are the schools' mains strengths?</li> <li>● What would you say are some of the school's biggest challenges?</li> <li>● What do you like most about the school?</li> </ul>	<p>Allows the participant an opportunity to settle into the interview, as well as providing some background to the school, so that we understand more about the context in which the Speech Bubbles programme is being delivered.</p>

<b>3. Engagement:</b>	<b>4 mins</b>
<p>How did your school become involved with the Speech Bubble programme?</p> <ul style="list-style-type: none"><li>● Why did your school decide to get involved in the programme?</li><li>● What initial expectations did you have for the programme?</li><li>● Were there any challenges you envisaged?</li></ul> <p>Can you tell me about your involvement with the Speech Bubbles programme?</p> <ul style="list-style-type: none"><li>● Can you describe any support that you, or other members of SLT, have provided to the programme?</li><li>● Can you describe any resources that you've made available to enable the programme to run?</li></ul> <p>Have you sat in on any of the Speech Bubbles sessions?</p> <p>If yes:</p> <ul style="list-style-type: none"><li>● Can you describe any sessions you've sat in on?</li><li>● What did you think of the session[s]?</li></ul> <p>[probe for further exploration of strength and suggested areas for improvement]</p> <p>If no:</p> <ul style="list-style-type: none"><li>● What do you understand about what the students have been doing in the Speech Bubbles sessions?</li></ul>	

4. Delivery experience	5 mins
<p>Can you tell me a little bit about the support that is normally offered for students in [intervention year] with communication and social skills difficulties at your school?</p> <ul style="list-style-type: none"> <li>● Have you continued to offer these activities alongside the Speech Bubbles programme?</li> <li>● What did you hope the Speech Bubbles programme would add?</li> </ul> <p>Could you tell me a little more about the school's pre-existing Drama Provision?</p> <ul style="list-style-type: none"> <li>● What kind of drama activities would pupils in [intervention year] typically receive?</li> <li>● How do you see the Speech Bubbles programme fitting in with other drama activities in the school? [probe for similarity/difference]</li> </ul> <p>Thinking specifically about the delivery of the programme, how have you found having the Speech Bubbles programme running in your school?</p> <ul style="list-style-type: none"> <li>● What has worked well about the delivery of the programme?</li> <li>● What have been the main challenges to delivering the programme? [e.g. working with drama practitioner/external agencies/staff buy-in, facilities, resources]</li> <li>● What have you done to try and overcome these challenges?</li> <li>● How have you found working with the Drama Practitioner?</li> <li>● How has the school found accommodating the weekly Speech Bubbles session into the school timetable?</li> <li>● Knowing what you know now, would you approach anything about the programme implementation differently?</li> </ul>	<p>This section will focus on understanding the perceived quality of the intervention, as well as the experience of the programme's delivery, including barriers and facilitators to delivery. We will also try and disentangle what is 'business as normal' within the school in terms of speech and language support and drama, and whether this has changed as a result of the Speech Bubbles programme.</p>

5. Mechanisms	4 mins
<p>What do you think the impact of the Speech Bubbles programme has been on targeted students?</p> <p>[probe for effect on classroom engagement and learning, and speech and communication. e.g.Can you describe any effect of the programme on pupils' communication skills?]</p> <ul style="list-style-type: none"> <li>• What was it about the programme that you think led to the effects you have described?</li> <li>• Are you aware of any students that the programme has been particularly helpful/unhelpful for?</li> </ul> <p>What effect, if any, do you think the Speech Bubbles has had for staff most closely involve with the programme?</p> <p>Teaching Assistant:</p> <ul style="list-style-type: none"> <li>• ...during the sessions?</li> <li>• ...back in the classroom?</li> <li>• What do you think it is about the Speech Bubbles programme that has led to these changes?</li> </ul> <p>Class Teacher:</p> <ul style="list-style-type: none"> <li>• ...in the classroom?</li> </ul> <p>Can you describe any effect of the programme more broadly within the school?</p> <ul style="list-style-type: none"> <li>• Can you describe any effect of the programme on other pupils within the school?</li> <li>• Can you describe any effect of the programme on other teachers within the school?</li> <li>• Has [intervention class] doing Speech Bubbles affected drama/speech and language provision within the school more generally? If yes, how?</li> </ul>	<p>To explore the SLT member's perception of the positive and negative impact of the programme, particularly for pupils, and the mechanisms that brought about any change identified.</p>
6. Close	2 mins
<p>Overall, would you recommend the programme to other schools?</p> <ul style="list-style-type: none"> <li>• Why/Why not?</li> </ul> <p>If you had a magic wand, what one thing would you change about Speech Bubbles and why?</p> <p>Is there anything else that you'd like to comment on before we close?</p>	<p>Thank you and close</p>

Thank the interviewee for their time and reassure them of the confidentiality of their responses, as explained at the beginning of the interview.	
---------------------------------------------------------------------------------------------------------------------------------------------------	--

## N4. Drama practitioner interview guide

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

Main objective	Purpose of section	Guide timings
1. Introductions	Explains the purpose and ground rules for the interview.	3 mins
2. Background context	Allows the participant an opportunity to settle into the interview, as well as providing some background to the school and the context in which the Speech Bubbles programme is being delivered.	5 mins
3. Delivery experience	To understand the perceived quality of the intervention, as well as experiences of the programme's delivery, including barriers and facilitators to delivery.	5 mins
4. Pupil engagement and programme mechanism	To explore the drama practitioner's perception of pupils' engagement in the sessions and the positive and negative impact of the programme, together with the mechanisms that brought about any impact identified.	5 mins
5. Training	To understand drama practitioner's experience of the training they received to deliver the programme.	2 mins
6. School Support	To understand the support they have received to deliver the programme, including from the school's SLT and class teachers.	3 mins
7. Close	Thank you and close.	2 mins

Observation of SB session		
8. Reflection following observation of SB session	Where possible, this will be an opportunity to follow-up on any areas of interest arising from the observation.	5 mins

### Topic guide

1. Introductions	5 mins
<p>Introduction:</p> <ul style="list-style-type: none"> <li>● Introduce yourself</li> <li>● Introduce BIT – explain that we are independently evaluating the Speech Bubbles programme, which is one of five programmes that are part of the Cultural Learning programme that is jointly funded by the Education Endowment Foundation and Royal Society of Arts.</li> </ul> <p>Aims of this interview:</p> <p>We are here to learn more about how the Speech Bubbles programme has worked in your class. We're interested in what involvement you have had with the programme and what has helped the programme to work, and what the challenges have been. We'd also like to understand any impact the programme has had on your school, particularly pupils in your class.</p> <p>This interview:</p> <ul style="list-style-type: none"> <li>● Should take no more than 30 minutes</li> <li>● Stress that you want to understand the world from your (the respondent's) point of view. No answers are right or wrong – and we are not here to judge the decisions made or views held by the interviewee.</li> </ul> <p>Anonymity and privacy:</p> <ul style="list-style-type: none"> <li>● All information gathered will be in strict confidence, unless there are concerns about safeguarding, and no-one will be named in any subsequent write-up of this research.</li> <li>● Explain that if at any point they feel uncomfortable or prefer not to answer a specific question they can just say so.</li> <li>● Explain that it is their choice whether they take part in the interview and they can end the interview at any point, without giving a reason.</li> </ul> <p>Recording:</p>	<p>Orientates respondent and gets them prepared to take part in the discussion.</p> <p>Outlines the 'rules' of the interview.</p>

<ul style="list-style-type: none"> <li>● Explain that recording enables us to have an accurate record of what was said, which can be typed up for analysis alongside other interviews. We may also use quotes from this interview, but these will be included in a way that means no individual or school is identifiable.</li> <li>● Check if they have any questions about the interview. If they are happy to go ahead, obtain verbal permission to digitally record and take notes (written permission should already have been obtained).</li> <li>● Once you have consent, start the voice recorder.</li> <li>● State interview number/participant ID.</li> </ul>	
<p><b>2. Background context</b></p>	<p><b>5 mins</b></p>
<p>How did you become involved with the Speech Bubbles project?</p> <ul style="list-style-type: none"> <li>● What is the programme trying to achieve?</li> <li>● Have you previously been involved in any similar programmes?</li> <li>● How does Speech Bubbles compare to these programmes?</li> </ul> <p>How would you describe this school?</p> <ul style="list-style-type: none"> <li>● Are there any particular challenges you think it faces?</li> </ul>	<p>Allows the participant an opportunity to settle into the interview, as well as providing some background to the school and the context in which the Speech Bubbles programme was delivered.</p>
<p><b>3. Delivery Experience</b></p>	<p><b>5 mins</b></p>
<p>Could you tell me about your role in delivering the Speech Bubbles programme/ sessions at [school name]?</p> <ul style="list-style-type: none"> <li>● What does each session typically consist of?</li> <li>● What has worked well about delivering the Speech Bubbles programme at this school?</li> <li>● What have the main challenges been?</li> <li>● How have you tackled them?</li> <li>● How suitable have the schools facilities been for the Speech Bubbles sessions?</li> <li>● Knowing what you know now, would you approach anything about the programme implementation within [case study school] differently?</li> </ul> <p>Can you tell me a little bit about working with the Teaching Assistant?</p> <ul style="list-style-type: none"> <li>● What is their typical role within the session?</li> <li>● How did you support them to engage with Speech Bubbles?</li> <li>● Did you experience any challenges in working with them?</li> </ul>	<p>This section will focus on understanding the perceived quality of the intervention, as well as experiences of the programme's delivery, including barriers and facilitators to delivery.</p>

<ul style="list-style-type: none"> <li>● If so, how did you overcome these challenges?</li> <li>● Have you seen any changes in their participation over time?</li> <li>● If so, what do you think this is down to?</li> </ul>	
<p><b>3. Pupil Engagement and Programme Mechanisms</b></p>	<p><b>10 mins</b></p>
<p>How have children in your sessions found the Speech Bubbles sessions?</p> <ul style="list-style-type: none"> <li>● Could you describe the pupils who have got the most out of the sessions? Why do you think this is?</li> <li>● Could you describe any children who have struggled more with the sessions? Why do you think this is?</li> </ul> <p>What effects have the Speech Bubbles sessions had for the pupils in your sessions?</p> <ul style="list-style-type: none"> <li>● Could you describe any changes you've noticed in your pupils? [if not mentioned, probe for effect on speech and communication and social skills e.g. Can you describe any effect of the programme on pupil's communication skills?]</li> <li>● Can you identify one aspect of the programme that has helped the most/anything that's been unhelpful?</li> </ul> <p>What effect do you think taking part in the Speech Bubbles programme has on the Teaching Assistant?</p> <p>[probe for professional development, confidence, role in the classroom, ownership, relationship with students]</p> <ul style="list-style-type: none"> <li>● When did you notice these changes?</li> <li>● What was it about the programme that you think led to these changes?</li> </ul> <p>How appropriate were the pupil referrals the school made to the programme?</p> <ul style="list-style-type: none"> <li>● Knowing what you know now, are there any changes you would advise the school make to the pupil referral process?</li> </ul> <p>Do you find yourself adapting the Speech Bubbles programme to support the needs of different schools?</p> <ul style="list-style-type: none"> <li>● Could you tell me a little more about this?</li> </ul>	<p>To explore the drama practitioner's perception of pupils' engagement in the sessions and the positive and negative impact of the programme, together with the mechanisms that brought about any impact identified.</p>

<ul style="list-style-type: none"> <li>● Did you use any particular physical resources as part of your Speech Bubbles sessions?</li> </ul> <p>If you could change one thing about Speech Bubbles what would it be and why?</p>	
<p><b>5. Training</b></p>	<p><b>2 mins</b></p>
<p>Can you tell me about the training sessions that you attended as part of the Speech Bubbles programme?</p> <ul style="list-style-type: none"> <li>● What did you cover in the training sessions?</li> <li>● Were they helpful?</li> <li>● Anything they might have overlooked/ with the benefit of hindsight could be changed?</li> </ul>	<p>To understand drama practitioner's experience of the training they received to deliver the programme,</p>

<b>6. School support for Speech Bubbles in the school</b>	<b>3 mins</b>
<p>How have the SLT responded to the Speech Bubbles programme at [case study school]?</p> <ul style="list-style-type: none"> <li>• Can you describe any support you've received from SLT to deliver the programme?</li> <li>• Can you describe any resources they've made available to deliver the programme?</li> </ul> <p>How has the class teacher responded to the Speech Bubbles programme?</p> <ul style="list-style-type: none"> <li>• Can you describe any support you've received from the class teacher to deliver the programme?</li> </ul> <p>How has the programme been perceived by parents of participating students?</p>	<p>To understand the support they have received to deliver the programme, including from the school's SLT and class teachers.</p>
<b>7. Close</b>	<b>2 mins</b>
<p>Is there anything else that you'd like to comment on before we close?</p> <p>Thank the interviewee for their time and reassure them of the confidentiality of their responses, as explained at the beginning of the interview.</p>	<p>Thank you and close</p>

<b>8. Reflections on observations</b>	<b>5 mins</b>
<p><i>Below are some exemplar questions, however, the interviewer may want to add some additional questions based on their observations. Probe for their perception of student engagement; implementer support; and mechanisms.</i></p> <p><i>Use the 'student evaluation sheet' produced by TAs and TPs after the session as a prompt.</i></p> <p>How did you find the session I have just observed?</p> <ul style="list-style-type: none"> <li>• What did you understand your role to be within the session?</li> <li>• What was the Teaching Assistant's role within that session?</li> <li>• What do you think went well?</li> <li>• What do you think were the challenges?</li> </ul> <p><i>[Use the student evaluation sheet as a prompt]</i></p> <ul style="list-style-type: none"> <li>• Which students engaged particularly well? How/Why?</li> <li>• Which students engaged less well? How/Why?</li> <li>• How did that session compare to previous sessions?</li> </ul>	<p>Opportunity to follow-up on any areas of interest arising from the observation.</p>

- |                                                                                                                                                                                                                                                                                      |  |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|
| <ul style="list-style-type: none"> <li>● Could you describe any ways that the sessions have changed over time?</li> <li>● What do you hope to do in future Speech Bubbles Sessions?</li> <li>● If you could go back and do that session again, what would do differently?</li> </ul> |  |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|

## N5. Guide for informal discussions with pupils

Short discussions, lasting around five minutes will be conducted with pupils happy to speak to a researcher. These will take place in the Speech Bubbles session. The researcher(s) will take field notes and therefore the conversations will not be audio recorded.

### **Aims of discussions with pupils**

To understand whether pupils are engaged with Speech Bubbles.

### **Before conducting discussions**

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

### **Discussion guide questions**

Tell me about your Speech Bubbles sessions

- What happens in your Speech Bubbles sessions?
- What do you get to do in your Speech Bubbles sessions?

Can you tell me one thing you like about your Speech Bubbles sessions?

Can you tell me one thing you don't like so much about your Speech Bubbles sessions?

- If you could change one thing about your Speech Bubbles sessions, what would it be?

Is there anything else you do in school that's like Speech Bubbles sessions or is Speech Bubbles very different from everything else?

- Like: Drama/ storytelling/ PE - tell me about them
- Different: TA/ Drama Practitioner/ teacher roles

Is there anything else you want to tell me about your Speech Bubbles session today?

Thank you for talking to me today.

## Appendix O: Observation pro forma

<p><b>Date and time:</b>  <b>Name of school:</b>  <b>Number of pupils present:</b>  <b>Name of observer:</b>  <b>Other adults present:</b></p>	
<b>Notes</b>	
<b>Description of setting and space used</b>	
<b>How is the session introduced, i.e. what are the aims and expectations?</b>	
<b>What activities are carried out as part of the session?</b>	
<b>How confident is the Drama Practitioner delivering the session?</b>	
<b>How confident is the Teaching Assistant delivering the session? (who leads the session and how are tasks shared?)</b>	
<b>To what extent are children engaged in the session?</b>	
<b>What strategies are used to manage pupils behaviour?</b>	
<b>To what extent do children seem able to understand and follow the session? What difficulties do children who struggle seem to have? How are they supported?</b>	
<b>What strategies are used to account for the needs and abilities of different pupils?</b>	
<b>Other</b>	

## Appendix P: Full survey content and results

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

1.1 Did you co-deliver the Speech Bubbles programme at your school?	
Response	Number (%)
Yes	21 (100%)
No	0 (0%)
Don't know	0 (0%)

1.2 What is your role at the school?	
Response	Number (%)
Teaching Assistant	17 (81%)
Other (please specify):	
- LEAD Speech & Language	1 (5%)
- Learning Support Assistant	1 (5%)
- Nursery Nurse	1 (5%)
- PE Lead	1 (5%)
- Performance Specialist	1 (5%)

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

2.1 Did you attend the induction training for Speech Bubbles in September?	
Response	Number (%)
Yes	19 (90%)
No	1 (5%)
Don't know	1 (5%)

N.B. Questions 2.2-2.6 were not asked to participants who responded 'no' to attending the induction training.

2.2 To what extent do you agree or disagree with the following statement: After the induction session, I understood the purpose of Speech Bubbles	
Response	Number (%)
Agree	10 (53%)

Strongly agree	9 (47%)
Neither agree nor disagree	0 (0%)
Disagree	0 (0%)
Strongly disagree	0 (0%)
Don't know	0 (0%)

**2.3 To what extent do you agree or disagree with the following statement:** After the induction session, I could co-deliver basic group exercises (e.g. story square)

Response	Number (%)
Agree	11 (58%)
Strongly agree	8 (42%)
Neither agree nor disagree	0 (0%)
Disagree	0 (0%)
Strongly disagree	0 (0%)
Don't know	0 (0%)

**2.4 To what extent do you agree or disagree with the following statement:** After the induction session, I could keep children engaged in the exercises

Response	Number (%)
Agree	12 (63%)
Strongly agree	7 (37%)
Neither agree nor disagree	0 (0%)
Disagree	0 (0%)
Strongly disagree	0 (0%)
Don't know	0 (0%)

**2.5 To what extent do you agree or disagree with the following statement:** After the induction session, I had all the materials required to co-deliver Speech Bubbles

Response	Number (%)
Agree	9 (47%)
Strongly agree	6 (32%)
Neither agree nor disagree	4 (21%)
Disagree	0 (0%)
Strongly disagree	0 (0%)
Don't know	0 (0%)

**2.6 To what extent do you agree or disagree with the following statement: After the induction session, I had the space required to co-deliver Speech Bubbles**

Response	Number (%)
Strongly agree	9 (47%)
Agree	9 (47%)
Neither agree nor disagree	1 (5%)
Disagree	0 (0%)
Strongly disagree	0 (0%)
Don't know	0 (0%)

**2.7 What were your reasons for not attending the induction in September?**

Response	Number (%)
I could not travel the distance required	1 (5%)
I had personal commitments	1 (5%)
Other (please specify): - Recently started covering SB sessions	1 (5%)
I could not find cover	0 (0%)
I did not receive time off in lieu	0 (0%)
Cost of travel and/or subsistence	0 (0%)
I had competing school commitments	0 (0%)

I did not feel my attendance would be useful	0 (0%)
I did not know about the training	0 (0%)
I was unwell	0 (0%)
SLT were not supportive of me attending	0 (0%)

N.B. Total percentage is below 100% because only respondents who had not attended the induction completed this question.

### 3.1. Did you attend the Collaborative Learning half-day session in January?

Response	Number (%)
Yes	18 (86%)
No	3 (14%)
Don't know	0 (0%)

N.B. Questions 3.2-3.4 were not asked to participants who responded 'no' to attending the collaborative learning event

### 3.2 To what extent do you agree or disagree with the following statements: The Collaborative Learning event was an opportunity to share my experiences of Speech Bubbles with others

Response	Number (%)
Strongly agree	11 (61%)
Agree	7 (39%)
Neither agree nor disagree	0 (0%)
Disagree	0 (0%)
Strongly disagree	0 (0%)
Don't know	0 (0%)

### 3.3 To what extent do you agree or disagree with the following statements: The Collaborative Learning event allowed me to problem-solve any challenges I had experienced when co-delivering Speech Bubbles

Response	Number (%)
Agree	9 (50%)
Strongly agree	8 (44%)

Neither agree nor disagree	1 (6%)
Disagree	0 (0%)
Strongly disagree	0 (0%)
Don't know	0 (0%)

**3.4 To what extent do you agree or disagree with the following statements:** Following the Collaborative Learning event, I had a clear plan about how to overcome any challenges I had experienced

Response	Number (%)
Agree	11 (61%)
Strongly agree	5 (28%)
Neither agree nor disagree	1 (6%)
Disagree	1 (6%)
Strongly disagree	0 (0%)
Don't know	0 (0%)

**3.5. What were your reasons for not attending the Collaborative Learning event?**

Response	Number (%)
I had personal commitments	1 (5%)
I did not know about the training	1 (5%)
Other: Recently started covering SB sessions	1 (5%)
I could not find cover	0 (0%)
I could not travel the distance required	0 (0%)
I did not receive time off in lieu	0 (0%)
Cost of travel and/or subsistence	0 (0%)
I had competing school commitments	0 (0%)
I did not feel my attendance would be useful	0 (0%)
I was unwell	0 (0%)

SLT were not supportive of me attending	0 (0%)
-----------------------------------------	--------

N.B. Total percentage is below 100% because only respondents who had not attended the Collaborative Learning sessions completed it.

**3.6. After you attended the Induction and/or Collaborative Learning sessions, how prepared did you feel to co-deliver the Speech Bubbles sessions?**

Response	Number (%)
I felt very prepared	14 (78%)
I felt somewhat prepared	4 (22%)
I felt somewhat unprepared	0 (0%)
I did not feel prepared at all	0 (0%)
Don't know	0 (0%)

**4.1 To what extent do you agree or disagree with the following statement: I understand the purpose of Speech Bubbles**

Response	Number (%)
Strongly agree	13 (62%)
Agree	7 (33%)
Neither agree nor disagree	1 (5%)
Disagree	0 (0%)
Strongly disagree	0 (0%)
Don't know	0 (0%)

**4.2 To what extent do you agree or disagree with the following statement: I can co-deliver basic group exercises (e.g. story square)**

Response	Number (%)
Strongly agree	13 (62%)
Agree	7 (33%)
Neither agree nor disagree	1 (5%)
Disagree	0 (0%)

Strongly disagree	0 (0%)
Don't know	0 (0%)

**4.3 To what extent do you agree or disagree with the following statement: I can keep children engaged in the exercises**

Response	Number (%)
Agree	12 (57%)
Strongly agree	8 (38%)
Neither agree nor disagree	1 (5%)
Disagree	0 (0%)
Strongly disagree	0 (0%)
Don't know	0 (0%)

**4.4 To what extent do you agree or disagree with the following statement: I have all the materials required to co-deliver Speech Bubbles**

Response	Number (%)
Agree	12 (57%)
Strongly agree	8 (38%)
Neither agree nor disagree	1 (5%)
Disagree	0 (0%)
Strongly disagree	0 (0%)
Don't know	0 (0%)

**4.5 To what extent do you agree or disagree with the following statement: I have the space required to co-deliver Speech Bubbles**

Response	Number (%)
Strongly agree	10 (48%)
Agree	10 (48%)

Neither agree nor disagree	1 (5%)
Disagree	0 (0%)
Strongly disagree	0 (0%)
Don't know	0 (0%)

**5.1. To what extent do you think the Speech Bubbles programme is engaging for the children who it is aimed at?**

Response	Number (%)
Very engaging	12 (57%)
Somewhat engaging	8 (38%)
Neither engaging nor unengaging	1 (5%)
Somewhat unengaging	0 (0%)
Very unengaging	0 (0%)
Don't know	0 (0%)

**6.1 Thinking about the children who took part in Speech Bubbles, what kind of impact, if any, do you think Speech Bubbles has had on pupils' drama skills?**

Response	Number (%)
Very positive impact	12 (57%)
Somewhat positive impact	8 (38%)
Neither positive nor negative impact	1 (5%)
Somewhat negative impact	0 (0%)
Very negative impact	0 (0%)
Don't know	0 (0%)

**6.2 Thinking about the children who took part in Speech Bubbles, what kind of impact, if any, do you think Speech Bubbles has had on pupils' behaviour?**

Response	Number (%)
----------	------------

Somewhat positive impact	9 (43%)
Neither positive nor negative impact	6 (29%)
Very positive impact	5 (24%)
Somewhat negative impact	1 (5%)
Very negative impact	0 (0%)
Don't know	0 (0%)

**6.3 Thinking about the children who took part in Speech Bubbles, what kind of impact, if any, do you think Speech Bubbles has had on pupils' reading?**

Response	Number (%)
Neither positive nor negative impact	7 (33%)
Don't know	6 (29%)
Somewhat positive impact	5 (24%)
Very positive impact	3 (14%)
Somewhat negative impact	0 (0%)
Very negative impact	0 (0%)

**6.4 Thinking about the children who took part in Speech Bubbles, what kind of impact, if any, do you think Speech Bubbles has had on pupils' speech and language?**

Response	Number (%)
Very positive impact	9 (43%)
Somewhat positive impact	8 (38%)
Neither positive nor negative impact	4 (19%)
Somewhat negative impact	0 (0%)
Very negative impact	0 (0%)
Don't know	0 (0%)

**6.5 Thinking about the children who took part in Speech Bubbles, what kind of impact,**

**if any, do you think Speech Bubbles has had on pupils' communication?**

<b>Response</b>	<b>Number (%)</b>
Very positive impact	13 (62%)
Somewhat positive impact	5 (24%)
Neither positive nor negative impact	3 (14%)
Somewhat negative impact	0 (0%)
Very negative impact	0 (0%)
Don't know	0 (0%)

**6.6 Thinking about the children who took part in Speech Bubbles, what kind of impact, if any, do you think Speech Bubbles has had on pupils' social skills?**

<b>Response</b>	<b>Number (%)</b>
Very positive impact	12 (57%)
Somewhat positive impact	8 (38%)
Don't know	1 (5%)
Neither positive nor negative impact	0 (0%)
Somewhat negative impact	0 (0%)
Very negative impact	0 (0%)

**6.7 Thinking about the children who took part in Speech Bubbles, what kind of impact, if any, do you think Speech Bubbles has had on pupils' creativity?**

<b>Response</b>	<b>Number (%)</b>
Very positive impact	10 (48%)
Somewhat positive impact	8 (38%)
Neither positive nor negative impact	2 (10%)
Don't know	1 (5%)
Somewhat negative impact	0 (0%)
Very negative impact	0 (0%)

**6.8 Thinking about the children who took part in Speech Bubbles, what kind of impact, if any, do you think Speech Bubbles has had on pupils' engagement?**

<b>Response</b>	<b>Number (%)</b>
Somewhat positive impact	10 (48%)
Very positive impact	8 (38%)
Neither positive nor negative impact	3 (14%)
Somewhat negative impact	0 (0%)
Very negative impact	0 (0%)
Don't know	0 (0%)

**6.9 Thinking about the children who took part in Speech Bubbles, what kind of impact, if any, do you think Speech Bubbles has had on pupils' confidence?**

<b>Response</b>	<b>Number (%)</b>
Very positive impact	13 (62%)
Somewhat positive impact	6 (29%)
Neither positive nor negative impact	1 (5%)
Don't know	1 (5%)
Somewhat negative impact	0 (0%)
Very negative impact	0 (0%)

**7.1 Outside the Speech Bubbles programme, what kind of support is offered to pupils with speech, language and communication challenges in your school?**

<b>Response</b>	<b>Number (%) answered 'Yes'</b>
In-class adult support	19 (90%)
Out-of-class adult support	15 (71%)
Adaptation of class teaching and materials	11 (52%)
Other (please specify):	
- Speech and language therapist	3 (14%)
- Various intervention groups	2 (10%)

- Staff training from Ed Psych	1 (5%)
- Specialised SEN teacher	1 (5%)
- Support from external agencies	1 (5%)
No support is offered	0 (0%)

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

**8.1 You worked with a drama specialist to co-deliver Speech Bubbles. How satisfied were you with their delivery of Speech Bubbles sessions?**

Response	Number (%)
Very satisfied	15 (71%)
Somewhat satisfied	5 (24%)
Neither satisfied nor dissatisfied	1 (5%)
Somewhat dissatisfied	0 (0%)
Very dissatisfied	0 (0%)
Don't know	0 (0%)

**9.1 What impact, if any, do you think the Speech Bubbles programme had on the following: My relationship with pupils taking part in Speech Bubbles**

Response	Number (%)
Very positive impact	12 (57%)
Somewhat positive impact	8 (38%)
Don't know	1 (5%)
No impact	0 (0%)
Somewhat negative impact	0 (0%)
Very negative impact	0 (0%)

**9.2 What impact, if any, do you think the Speech Bubbles programme had on the following: My understanding of how children learn**

Response	Number (%)
Somewhat positive impact	10 (48%)

Very positive impact	8 (38%)
No impact	2 (10%)
Don't know	1 (5%)
Somewhat negative impact	0 (0%)
Very negative impact	0 (0%)

**9.3 What impact, if any, do you think the Speech Bubbles programme had on the following: My confidence in the classroom**

Response	Number (%)
Very positive impact	8 (38%)
Somewhat positive impact	8 (38%)
No impact	4 (19%)
Don't know	1 (5%)
Somewhat negative impact	0 (0%)
Very negative impact	0 (0%)

**9.4 What impact, if any, do you think the Speech Bubbles programme had on the following: My job satisfaction**

Response	Number (%)
Very positive impact	9 (43%)
Somewhat positive impact	7 (33%)
No impact	5 (24%)
Don't know	0 (0%)
Somewhat negative impact	0 (0%)
Very negative impact	0 (0%)

**9.5 What impact, if any, do you think the Speech Bubbles programme had on the following: My relationship with the class teachers**

Response	Number (%)
No impact	8 (38%)
Somewhat positive impact	7 (33%)
Very positive impact	6 (29%)
Don't know	0 (0%)
Somewhat negative impact	0 (0%)
Very negative impact	0 (0%)

**9.6 What impact, if any, do you think the Speech Bubbles programme had on the following: My relationship with SLT (Senior Leadership Team)**

Response	Number (%)
No impact	9 (43%)
Very positive impact	7 (33%)
Somewhat positive impact	5 (24%)
Don't know	0 (0%)
Somewhat negative impact	0 (0%)
Very negative impact	0 (0%)

**10.1. During this school year, how many Speech Bubbles sessions have taken place?**

Response	Number (%)
20+	18 (86%)
16-20	3 (14%)
5 or less	0 (0%)
6-10	0 (0%)
10-15	0 (0%)
None	0 (0%)
Don't know	0 (0%)

**10.2 What were the main challenges you experienced that prevented you from running the drama sessions once a week?**

<b>Response</b>	<b>Number (%) answered 'Yes'</b>
It was hard to prioritise Speech Bubbles relative to other learning objectives	1 (5%)
Other (please specify):	
- Whole school training	1 (5%)
- None	1 (5%)
The children did not enjoy the sessions	0 (0%)
SLT were not supportive	0 (0%)
The children did not want to engage in the sessions	0 (0%)
I did not have the materials required to co-deliver the sessions	0 (0%)
Time is too limited in the curriculum	0 (0%)
The right space was not available	0 (0%)

N.B. Total percentage is below 100% because only respondents who reported having done less than 20 Speech Bubbles sessions completed it.

You may re-use this document/publication (not including logos) free of charge in any format or medium, under the terms of the Open Government Licence v3.0.

To view this licence, visit <https://nationalarchives.gov.uk/doc/open-government-licence/version/3> or email: [psi@nationalarchives.gov.uk](mailto:psi@nationalarchives.gov.uk)

Where we have identified any third-party copyright information you will need to obtain permission from the copyright holders concerned. The views expressed in this report are the authors' and do not necessarily reflect those of the Department for Education.

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



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

<https://educationendowmentfoundation.org.uk>

 [@EducEndowFoundn](https://twitter.com/EducEndowFoundn)

 [Facebook.com/EducEndowFoundn](https://www.facebook.com/EducEndowFoundn)