

**Independent evaluation of Talk with Tales for
Children (TWiTCH): A two-armed cluster
randomised controlled trial**



Evaluation Protocol

Evaluator (institution): Durham University (with the University of Manchester and Navigator Insight)

Principal investigator: Vic Menzies

Template last updated: August 2025

Evaluation summary

Project title	Independent evaluation of Talk with Tales for Children (TWiTCH): A two-armed cluster randomised controlled efficacy trial
Developer <i>(Institution)</i>	Sheffield Hallam University
Evaluator <i>(Institution)</i>	Durham University with the University of Manchester and Navigator Insight
Principal investigator(s)	Vic Menzies
Protocol author(s)	Vic Menzies, Akansha Singh, Johny Daniel, Katie Morris, Xiaofei Qi, Kelly Burgoyne
Trial design	Two-arm cluster randomised controlled trial with random allocation at setting level and business as usual control group
Trial type	Efficacy
Pupil age range and Key stage	3-4 years, EYFS
Number of settings <i>(at design stage)</i>	132
Number of pupils <i>(at design stage)</i>	1320
Primary outcome measure and source	Children's holistic language skills (LanguageScreen assessment)
Secondary outcome measure and source	1) Children's oral narrative language (Renfrew Bus Story Test) 2) School readiness (Early Years Foundation Stage Profile at end of Reception) 3) Practitioner shared book reading behaviours (bespoke coding framework used to analyse recordings of small group story time)

Protocol version history

Version	Date	Reason for revision
1.0 [original]	27/03/2026	N/A

Contents

Study rationale and background.....	5
Intervention	9
Impact evaluation design	22
Implementation and process evaluation (IPE) design.....	43
Cost evaluation design	57
Ethics and registration.....	57
Data protection	58
Personnel	60
Risks.....	62
Timeline	65
References	67
Appendix A: Changes since the previous EEF evaluation.....	74
Appendix B: Theory of Change Assumptions Logs.....	75
Appendix C: Audio Recording Coding Tool.....	82

Tables

Table 1. PLANIT Table describing TWITCH Programme.....	9
Table 2: Trial design.....	23
Table 3. Sample size calculations	34
Table 4. Expected attrition	34
Table 5. Exploring MDES when varying assumptions	36
Table 6. TWITCH Compliance Descriptor	40
Table 7. Implementation and Process Evaluation Research Methods	45
Table 8. Risk management approach	62
Table 9. Timeline.....	65

Figures

Figure 1. TWiTCH Logic Model.	21
------------------------------------	----

Appendix tables

Appendix Table 1: Changes since the previous evaluation	74
Appendix table 2. Theory of change contextual assumptions	75
Appendix table 3. Theory of change causal assumptions	77

Study rationale and background

Theoretical and scientific background

Importance of Early Language for Later Attainment

Early language and communication skills are recognised as foundational for children's learning and life outcomes. Research shows that language skills at age four predict academic achievement throughout primary school, with persistent difficulties linked to lower attainment (Snow et al., 1998; Eadie et al., 2021). The link between early outcomes can persist into secondary school and later academic outcomes (Law et al 2009). Children from disadvantaged backgrounds are more likely to have language delay at age five than their more advantaged peers (Law, Charlton & Asmussen, 2017), and these gaps tend to widen over time (Tymms et al., 2014; Merrell et al., 2014).

Sustained Shared Thinking and Dialogic Pedagogy

Large-scale UK evidence from the Effective Provision of Pre-School Education (EPPE) study, demonstrates that high-quality early years provision, particularly rich adult-child interactions, can positively influence oral language on school entry, as well as literacy, and academic outcomes through Key Stages 1 to 3 and at GCSE (Sylva et al., 2004; Sammons et al., 2004, 2011). A defining feature of effective provision is the quality of the adult-child verbal interaction and the specific use of dialogic pedagogy - using open-ended questions and discussion to support conceptual understanding and cognitive development (Alexander, 2017; Jay et al., 2017). In early years settings, these high-quality interactions are often conceptualised as *Sustained Shared Thinking (SST)* where adults and children co-construct meaning through dialogue and extending ideas across extended exchanges with genuine interest from the teacher in the child's idea and use of open questions. SST is a specific dialogic pedagogy used within early years contexts, however SST remains relatively rare in typical practice, amplifying the case for professional development (PD) that builds practitioners' capacity to engineer and sustain such dialogic interactions (Siraj-Blatchford & Manni, 2009).

Dialogic pedagogy (and within this, SST) seeks to promote classroom talk that is reflective, reasoned and reciprocal. The EEF's *Dialogic Teaching* efficacy trial, which trained Year 5 teachers to use dialogic strategies, reported improvements of an additional two months' progress in English and science compared to business-as-usual (Jay et al., 2017). Although based in primary school rather than early years, these findings support the theoretical premise that structured, high-quality dialogic talk can improve learning, and provide a relevant conceptual foundation for TWITCH, which adapts dialogic principles to story-time in nursery.

Shared book reading

Meta-analytic evidence demonstrates that interactive shared book reading can strengthen children's expressive and receptive language (Mol, Bus & de Jong, 2009; Flack, Field and horst, 2018). Mol, Bus & de Jong (2009) showed moderate positive effects on children's oral language skills across 31 experimental and quasi-experimental studies, highlighting the importance of

repeated reading and active child engagement during book sharing. Flack, Field & Horst (2018) further reported that the reading style, including use of dialogic prompts and repeated exposure of texts, significantly moderated vocabulary growth during storybook reading.

Meta-analyses of parent–child book sharing interventions have also reported small-to-moderate positive impacts on expressive and receptive language (Dowdall et al., 2020, Barone et al., 2019) with dialogic reading methods producing the most consistent language gains, though effects varied by socioeconomic context and some studies identified fade-out over time.

Together, this evidence indicates that shared book reading interactions can support language development, but that impact is likely to depend on implementation quality, practitioner skill, and opportunities for sustained, language-rich interaction—all core targets of TWiCH.

Reasoning language and P4C-influenced approaches

The Philosophy for Children (P4C) approach emphasises collaborative enquiry, conceptual talk and reasoning. The approach has been evaluated in primary schools through two EEF funded trials. In the first trial positive effects of approximately two additional months’ progress in reading and maths were found (Gorard, Siddiqui & See, 2015). However, the subsequent larger effectiveness trial (Lord, Dirie, Kettlewell & Styles, 2021) reported no average improvement in attainment, though teachers perceived gains in pupils’ communication and reasoning skills. Previous evaluations have not examined any P4C approaches in early years settings.

The TWiCH approach draws on P4C principles particularly for week 2 activities to supplement the other dialogic approaches. TWiCH incorporates the idea of concepts, a key part of the P4C approach. Conversations are framed around ideas that are *conceptual* (ideas worth talking about), *contested* (that invite different perspectives), and *connected* and *meaningful* to individuals’ life experience (www.thoughtful.org.uk). The choice-based language activities in week 2 draw on this theory to encourage better conversation and verbal fluency while encouraging deeper thinking. The TWiCH programme also incorporates high quality facilitation and scaffolding techniques based on Socratic traditions where the facilitator is not seen as an expert and uses open-ended questions to provoke deeper thinking. The Repeat, Check, Reason and Imagine (RCRI) interaction strategy that practitioners are trained to use, draws on the P4C facilitation approach. The incorporation of the concept-based conversation activities in week 2 and the use of the RCRI strategies is designed to extend the dialogic pedagogy (particularly in regards to distancing prompts used in the week 2 sessions) extending the opportunities for children to relate the story and the concepts in the stories, to their own experiences and encouraging the use of specific reasoning language and to promote higher order thinking.

Summary

The evidence suggests that raising the quality of adult–child talk through dialogic pedagogies, embedding interactive shared reading with repeated readings of the same story, and actively promoting reasoning language are promising methods for improving preschool language. However, effects are likely contingent on implementation fidelity, programme dosage, and

practitioner skill, strengthening the rationale for a professional development led model like TWiTCH that targets adult practice as the mechanism for improving child language outcomes.

The TWiTCH Programme

The Talk with Tales for Children (TWiTCH) programme integrates dialogic reading, opportunities for sustained shared thinking (SST), and the Philosophy for Children (P4C) elements of discussion around contested concepts and Socratic facilitation methods into a structured professional development programme for nursery staff. The programme is based around a structured cyclical programme for practitioners to implement in the classroom. It includes initial training and ongoing coaching for nursery staff delivered by early years and P4C experts, resources to deliver nine themed three-week cycles of activities based on traditional tales (with nurseries expected to deliver at least seven cycles), and additional support for a nursery Champion to lead delivery in the setting and to support other practitioners. TWiTCH activities are designed to be easily embedded within familiar practice in the Early Years classroom, replacing usual daily story-time activities. TWiTCH activities are delivered for 10 to 15 minutes each day in small groups of between 5 and 10 children which facilitate opportunities for dialogue. The programme aims to teach practitioners how to use story time as a daily opportunity for complex language interactions between children and adults, to support the use of teacher / child talk through Philosophy for Children inspired language games and practices and provide opportunities for Sustained Shared Thinking which encourage children's thinking and reasoning skills.

Previous Evaluation and delivery

The TWiTCH programme has undergone significant development and improvement since first being delivered in a small pilot of 5 nurseries which showed feasibility, acceptability and tentative promise on a small scale despite being interrupted by Covid-19 (Demissie, 2019).

A larger EEF-funded pilot evaluation, delivered through the DfE's Stronger Practice Hub initiative, worked with 25 nursery settings in Sheffield and surrounding areas in 2023/2024 to trial the TWiTCH approach across the course of a school year in settings with varying characteristics (Fitzfussell et al, 2025). The evaluation focused on the acceptability and feasibility of the programme across a variety of different types and sizes of settings, as well as looking for evidence of change to confidence, knowledge and practice for practitioners taking part in the programme.

Overall programme feasibility and acceptability was high, with almost all (22/25) participating settings implementing the programme through the 30-week delivery period and completing at least six of the-nine three-week TWiTCH story cycles (with 16/25 completing or almost completing the full nine cycles. This indicated that the core programme structure is deliverable across nursery settings with different characteristics. Implementation quality varied across the different aspects of the three-week cycle. The Week 1 shared reading activities were delivered with consistently high fidelity, whereas Week 2 reasoning activities required adaptation in many settings due to perceived task complexity and the mixed language levels of children in the setting. In week 3, settings consistently included TWiTCH book-related activities and resources for children to access and engage with, but there was variation in how children were identified for additional support and the delivery of additional support activities. These variations in delivery

indicated a need to provide guidance and resources for simplifying the week 2 activities and clearer guidance around providing additional support in week 3.

Through the pilot evaluation the programme showed evidence of promise. Practitioners reported increased confidence and use of dialogic and reasoning strategies, with audio-recorded observations showing greater use of open questions, extended pupil talk and language-scaffolding behaviours—evidence that the pilot successfully shifted practice in the intended direction. Practitioners also reported language improvements for children in the setting including improved vocabulary and oral communication skills as well as greater engagement with books and storytelling activities.

The pilot evaluation highlighted some practical challenges with implementing TWiTCH. Some settings had difficulties in facilitating the smaller group sizes due to the setting organisation and staffing issues which led to difficulties in using the TWiTCH techniques well due to needing to manage behaviour more and limited interaction opportunities with individual children. Children who attended parttime or who had irregular attendance patterns were not always able to access all the TWiTCH content due to the structure of the programme which included specific activities, prompts and questions related to the storybook for each of the five days sessions especially during week 1. This meant children who only attended at the end of the week would not be exposed to the prompts and discussion from earlier days which helped initial understanding of the stories. Training and support provided to Coaches was not always consistently delivered across coaches and settings suggested a need for clearer expectations, strengthened coach training, and quality-assurance systems.

Following the pilot evaluation, the TWiTCH programme was updated in line with the pilot report recommendations and delivered to a further cohort of 25 nursery settings in the Midlands. The programme now features:

- a strengthened coach training model which includes structured group training for all coaches together delivered over three half day sessions to support better understanding of the programme theory, clarification of coach expectations and more consistent delivery of setting training and support,
- updates to the practitioner handbook including updated week 2 activities and the provision of further guidance for adapting the games to the needs of the children involved,
- a change to the training for settings so that settings now implement a full cycle of TWiTCH in between the first and second training sessions so that coaches can provide support to settings based on their experiences and challenges with delivery of the first full cycle.
- revision to the classroom resources provided in line with practitioner and coach feedback of use.

Further details of how the TWiTCH programme has been changed from delivery in the pilot evaluation for the efficacy evaluation are included in Appendix A.

Although the pilot evaluation generated positive indications of practitioner change and practitioner-reported changes to children, it did not measure child level outcomes, and the design did not allow for attribution of causal effects. The combined findings indicate strong promise for the programme and emphasise the need now for a rigorous impact evaluation with an experimental design which is adequately powered to assess effects on children’s language. This efficacy trial aims to do this through the use of a cluster randomised controlled trial with a much larger number of settings and through a focus on child level language outcomes.

This evaluation is designed to provide a rigorous assessment of the causal impact of the TWiCH programme on children’s early language outcomes, while also generating evidence about implementation, fidelity, and mechanisms of change. Building on evidence of feasibility and promise from the previous EEF-funded pilot evaluation, this study adopts a two-arm cluster randomised controlled efficacy trial design, complemented by a comprehensive implementation and process evaluation (IPE).

The impact evaluation (IE) is powered to detect small but educationally meaningful effects on children’s holistic language skills, with a planned sample of 132 nursery settings (approximately 1,320 children) and a minimum detectable effect size of 0.18 for the primary outcome. Randomisation at setting level is necessary given that TWiCH is delivered as a whole-setting approach embedded within daily story-time practice, making individual randomisation infeasible due to contamination. The primary outcome—children’s holistic language skills measured using the LanguageScreen assessment—is closely aligned with the programme’s theory of change and captures the receptive, expressive and listening skills that TWiCH is explicitly designed to improve. Secondary outcomes include children’s oral narrative skills, practitioner shared reading behaviours, and longer-term school readiness outcomes derived from administrative data. This will enable exploration of outcomes at different levels of the theory of change from practitioner outputs to longer term impact.

The accompanying IPE is integral to the overall design. It is designed to assess the extent to which TWiCH is implemented with fidelity at scale, to document variation in delivery and adaptation across settings, and to further test key assumptions within the programme’s theory of change. The IPE will also examine usual practice in both intervention and control settings, explore contextual factors that may shape delivery, and identify mechanisms through which practitioner practice changes are expected to influence child outcomes. By triangulating qualitative and quantitative implementation data with impact estimates, the IPE will support interpretation of trial findings, inform the planned compliance and CACE analyses, and generate learning relevant to future scale-up.

Intervention

Table 1 below describes the TWiCH Programme in detail using the PLANIT framework. This description was initially written by the developer team and updated by the evaluation team after the set-up meetings.

Table 1. PLANIT Table describing TWiCH Programme

Purpose

Name of intervention

Talk With Tales for Children (TWiTCH)

Aims and objectives of intervention

The overall aim of the intervention is to train early years practitioners interaction styles during story time to improve children's receptive and expressive language.

The core components of the intervention are nine repeated three-week cycles based on different traditional story picture books with each three-week cycle encompassing:

- Developing knowledge and understanding of the story through repeated reading of the text and through use of dialogic reading strategies (Week 1)
- Language games to promote higher order thinking and language along with explicit use of Repeat, Check, Reason and Imagine strategies from the practitioner (Week 2)
- Embedding story in continuous provision and providing additional support for children who need additional input (Week 3).

TWiTCH activities are delivered multiple times during each week in small groups of up to 10 children. Practitioners are supported to deliver the programme through initial training adapted to the needs of the setting, and through ongoing coach support which includes termly meetings and observations of practice. A TWiTCH Champion in each setting receives additional training and attends networking events with other Champions to allow them to support and embed TWiTCH practices better in their setting.

It is expected that delivery of the three-week repeated reading cycle will motivate and engage children in TWiTCH sessions and lead to improved engagement with stories and vocabulary as well as improved children's confidence to articulate their thoughts and opinions. They are also expected to provide dialogic language and conversation opportunities between children and practitioners and increased incidences of sustained shared thinking.

Training on dialogic reading principles and ongoing support delivered by experienced and trained coaches will improve practitioners' skills in dialogic reading as well as practitioners' facilitative questioning using the RCRI (Repeat, Check, Reason, Imagine) strategies which will support the development of children's language skills.

The TWiTCH Champion will support the embedding and improvement of practice within the setting supported by additional training and networking opportunities.

Rationale for intervention

Concern about the language gap between advantaged and disadvantaged young children has been a driver for policy and interventions in the last decade. Language competency impacts across the curriculum when children enter school (Oxford University Press, 2018).

Language skills are essential to good levels of development¹ (Beitchman, Brownlie, & Bao, 2014) and are inextricably linked to outcomes in PSED (Kalland & Linnavalli, 2023). The development of language and self-regulation is bidirectional and better language skills promote pro-social skills and cooperative play (Tsomokos & Raviv, 2024). Additionally, TWiTCH aims to address the problem of decreasing levels of engagement with books and book sharing activities in young children’s lives. Concerns about the effect of screen time on language development are highlighting the importance of shared book reading in the early years (Massaroni et al., 2024).

Early years teachers know about the importance of high-quality interactions (Sylva et al 2004) but they have limited professional development opportunities and consequently lack skills and confidence in developing children’s language and communication (Purdon, 2016).

TWiTCH specifically draws from the following evidence:

- reading stories out loud on its own does not sufficiently improve language development (Noble et al., 2019)
- the quality and quantity of adult-child interactions in the early years have a significant impact on children’s language development (Siraj-Blatchford et al., 2002), and TWiTCH provides daily opportunities for high quality interactions
- Sustained Shared Thinking supports children’s thinking skills but is not always understood by practitioners and opportunities to use this form of dialogue are often missed (Sylva et al 2004). TWiTCH is focused on such dialogues and creating time for sustained shared thinking.
- Traditional stories and story books offer unique opportunities to have high quality interactions which explore deeper concepts and ideas (Haynes and Murris 2012)
- The use of dialogic pedagogy supports children’s self-expression skills and motivation (particularly for children on FSM) (Jay et al 2017), and TWiTCH trains practitioners to apply this pedagogy.
- Repeated reading enhances vocabulary and language development (Horst et al., 2011; Mihai, & Butera, 2019)

Learners

Participants

The programme targets all children aged 3-4 years in early years private, voluntary or independent (PVI) nursery settings and maintained settings in the year before they attend school. Targeted support is provided to individual ‘at need’ children in the third week of each cycle with practitioner’s deciding which children need additional support to engage with TWiTCH. Eligible settings are required to have at least two staff working with the pre-school cohort to enable testing of the TWiTCH Champion support model and a minimum of 15 children who attend at least 15 hours per week.

¹ In the Early Years Foundation Stage (EYFS), *Good Levels of Development* (GLD) refers to the summative assessment of children’s attainment by the end of the reception year. Children are considered to have achieved a good level of development if they have reached at least the ‘expected’ level in the EYFS areas of learning (DfE, 2026). For TWiTCH, GLDs relate specifically to the areas of language and communication (listening, attention and understanding, and speaking) and literacy (comprehension).

The programme aims to work with all practitioners who work with this age group but have specified a minimum of two staff should attend the training.

Actors

Intervention developer(s)

The intervention was developed by Dr Fufy Demissie and Professor Sally Pearse. Fufy is an early years lecturer at Sheffield Hallam University and an experienced trainer in dialogic pedagogy. She is CPD lead for the Early Years Community Research Centre (EYCRC) and has written several chapters and papers on early years pedagogy. She led on the EEF funded pilot for TWiATCH and the £42,500 SHINE funded project that preceded this evaluation project.

Professor Sally Pearse ran a 60-place community nursery rated as 'Outstanding' by Ofsted and an example of best practice for language and communication by National Strategies. At Sheffield Hallam, Sally is strategic lead for early years and Director of the EYCRC which has developed place-based approaches to supporting young children and their families.

The developer team will be involved in providing training and ongoing support to the coaches who will train and support settings in delivering TWiATCH. They will also provide direct online training to TWiATCH Champions and will lead the TWiATCH network meetings. If needed, the intervention developers may also serve as a coach to some settings.

Intervention instructor(s)

Coaches who are external to the developer's organisation will be employed by the developers as consultants to deliver training and support to nursery settings. These consultants will be trained coaches with expertise in early years/coaching and will be familiar with the Philosophy for Children (P4C) pedagogy. Some may have served as TWiATCH coaches in earlier trials of TWiATCH but this is not necessary.

All coaches will undertake three three-hour training sessions delivered by the developer team in May, June/July and September 2026 before the project starts. The first session will be delivered face-to-face, followed by two online sessions, covering the TWiATCH programme background, delivery, evaluation and ways to support settings. Coaches who are less familiar with the P4C pedagogy (who have not previously undertaken P4C training) will also attend a two-hour online training session covering the P4C approaches in TWiATCH (July 2026). While only coaches who have not previously undergone P4C training will be required to attend this training, the recording for the training and slides will be circulated to all coaches. Throughout programme delivery there will be three one-hour online meetings for coaches to get together with the developer team to share practice (in November 2026, February and May 2027), along with two focus groups to reflect on coaching experiences and improve the programme (in December 2026 and July 2027).

Intervention facilitator(s)

Setting practitioners/nursery teachers who work with 3–4-year-old children in their setting will deliver the intervention to pupils. There are no specific qualifications necessary for the delivery of TWiATCH. One practitioner in each setting will work as the TWiATCH Champion to lead and support delivery in their setting – this person should be experienced in supporting other practitioners or wanting to gain this experience.

Nature

Setting

The intervention will be conducted within private, voluntary or independent (PVI) nursery settings and within maintained state-funded nurseries (school nurseries and nursery only settings) in the Yorkshire and the Humber, East Midlands and North West areas of England.

The specific regions to be targeted are:

North West: Bolton, Bury, Knowsley, Liverpool, Manchester, Oldham, Rochdale, Salford, Stockport, Tameside, Trafford, Wigan, Cumbria, Lancashire, Warrington, Widnes, Sefton, St Helens

East Midlands: Derbyshire, Nottingham

Yorkshire and the Humber: Barnsley, Bradford, Calderdale, Doncaster, Kirklees, Leeds, Rotherham, Sheffield, Wakefield, York

These regions have been chosen due to proximity to existing coaches and established relationships with organisations who would support with recruitment in these areas. These regions include areas of high deprivation and priority will be given to settings in areas of disadvantage.

Intervention location

The intervention will take place in the nursery classroom or smaller spaces (accommodating up to 10 children) within the nursery for small group activities during weeks 1 and 2 of the cycle. The classroom and other areas of the nursery (e.g. outside space, sensory room) will be used for continuous provision activities.

Training and coaching for nursery practitioners will take place in the nursery setting.

Intervention timing and duration

The expected delivery of the TWiTCH repeated three-week story book cycles is for the delivery of nine cycles across 27 weeks between November 2026 and July 2027. In each three-week cycle the time expectations are:

- Week 1 - 10 minutes of TWiTCH story time activities every day. (SHU will reinforce this as part of Coach training.)
- Week 2 - 10 minutes of TWiTCH language game activities at least 3 times per week.
- Week 3 - the timing and during of activities will depend on the activities being provided as part of continuous provision. Activities may be specific one-off activities, or resources and toys provided for children's role play. The time required to target specific children will also depend on the needs and number of children that need additional support.

Instruction

Delivery – professional development

All practitioners

All practitioners working with 3–4-year-old children in the setting are invited to attend training with a minimum of two setting staff required to attend. Training for these practitioners involves:

- Watching a 10-minute pre-recorded 'Introduction to TWiTCH' video which outlines some of the TWiTCH programme background before attending the first in-person session. A shortened version of the video will be shown on Day 1 of training if practitioners have not seen the video before training or if practitioners would like to see it again.
- Attending two three-hour in-person training sessions held at each participating setting where the nursery is expected to complete the first TWiTCH cycle between these sessions. These sessions are delivered by the assigned TWiTCH Coach who is expected to model TWiTCH strategies and mentor staff on their use. The first training session covers the full TWiTCH three-week cycle and is tailored by the coach to the needs of the setting. The second session allows reflection and refinement of the nurseries implementation with the coach supporting how they can improve delivery going forward. Both training sessions provide opportunities for practitioners to role-play the activities and receive feedback from the coach and include time for planning and discussion.

TWiTCH Champions

TWiTCH Champions are a specific practitioner nominated by the setting to receive additional training and to support the delivery of TWiTCH across practitioners in the setting but will usually also deliver the TWiTCH programme themselves. If the TWiTCH Champion is a senior leader or nursery manager, they may not be able to deliver the programme themselves due to not working directly with the nursery class – they would still be expected to plan and support delivery of TWiTCH in other staff

The champion role includes:

- Building a relationship with and liaising with the coach
- Overseeing fidelity of TWiTCH delivery by liaising with other practitioners about how they are implementing the programme and providing support, suggestions, and ideas for better implementation
- Coordinating availability of storybooks and language activity resources
- Actively initiating peer reflection opportunities for colleagues.

Along with attending the coach-led in setting training with other practitioners from their setting, TWiTCH Champions also attend a two-hour online training session hosted by the developer team lead which aims to provide the TWiTCH champion with additional information about the role of the Champion, and deeper knowledge of the underlying theory and techniques used in TWiTCH. TWiTCH Champions are also invited to attend two online 30-minute TWiTCH champion network events in March and May 2027 during the delivery period. These sessions will bring TWiTCH champions from different settings to meet online together and are hosted by the developer team. These sessions are designed to encourage Champions to reflect on the delivery of TWiTCH in their setting and to ask for and receive feedback from both the developer team and the other TWiTCH champions.

Setting support meetings/visits

Coaches support setting delivery of TWiTCH through three coaching meetings with each setting (one per term). At least one of these meetings is conducted in the setting with coaches observing TWiTCH practice and providing feedback. The other meetings can take place either in the setting or online at the negotiation of the setting and the coach. These meetings are designed to encourage practitioners to reflect on how the programme is going and to help tackle any issues the setting is experiencing. Coaches are expected to act as a critical friend and support to the setting. Settings are encouraged to contact their coach if they need support outside of the planned meetings. Depending on the needs of the setting coaches may meet only with the TWiTCH Champion or with other practitioners in the setting.

Delivery – Modes and content

The main delivery method for the TWiTCH Programme is implementation of the three-week storybook cycle which is repeated nine times during the academic year. The TWiTCH activities are delivered by nursery practitioners to children in small groups of up to 10 children during week 1 and week 2 activities while week 3 activities can be self-initiated play, full class activities, small group or activities with individuals.

Activities for the three-week cycle involve:

Week 1: The story book for the three-week cycle is introduced to the children in small groups of ideally 5-8 children (but up to 10 is acceptable) and read to them at each daily session; drawing out sequence, concepts and characters. Each day's reading has a different focus: day one – overview of the story; day two – characters in the story; day three – sequencing; day four – listening to and enjoying the story; day five – prediction. There are prompts for practitioners about what to ask children in each session.

Week 2 – The focus for Week 2 is on language reasoning games that are based on the story and which are intended to develop children's language, enabling them to express complex ideas. Practitioners undertake 3 sessions of different choice-based language games over three days. The language games encourage the children to make a choice, give reasons and reflect on their reasons through language such as 'I agree because..' or 'I disagree because..' Practitioners are encouraged to use specific techniques to facilitate children's reflections including repeating and checking what the child says (REPEAT/CHECK), asking children to give reasons for their choices and what they say (REASONS) and Asking "What if..." questions (IMAGINE). Practitioners are encouraged to model reasoning language during the games using the rabbit puppet. It is expected that early Week 2 sessions will have more modelling behaviour but that children will develop confidence in taking part throughout the programme and be more involved in sessions later. Practitioners can also continue reading of the storybook to the whole class or group as required or requested by children.

Week 3: The story is used as the basis for activities offered across familiar daily activities as well as in story-time. These activities could be offered to the whole group or in smaller groups. Practitioners are asked to incorporate at least one planned story-related activity during this week. Activities could include, but are not limited to, story-scribing (where a child tells a story and a practitioner writes the story down word for word), one to one discussion, or playing alongside to narrate and recast children's language, provision of story-related role-playing

props. Practitioners are also encouraged to identify and intentionally focus their support and engagement on children with greater needs in language during Week 3 either one-to-one or in small groups. A template will be provided to help practitioners plan and track who they are giving extra support to. Children targeted for additional support could include those that missed other TWiTCH activities, those that are struggling to access the stories or children who need more input in language skills. Support for targeting children will be given in the Coach training.

Settings are provided with story books and materials to deliver nine cycles of the programme. It is not required for all practitioners in the setting to complete the nine cycles—just that the setting delivers each of the nine cycles. This could mean that one practitioner completes all of the cycles or that multiple practitioners complete all of the cycles with different groups or that practitioners share the delivery of the cycles. The TWiTCH Champion is expected to support the embedding of the TWiTCH practice across practitioners throughout the programme.

Delivery – Procedures

The TWiTCH programme should be embedded within settings' core curriculum delivery replacing settings' usual story-time practice.

Delivery - Materials

Nursery settings are provided with two setting resource packs to use when implementing the programme. This includes:

- 9 story books (so they can deliver the programme over nine three-week cycles)
- rabbit puppet (supports story reading & language games encouraging modelling of language)
- carrot plush toy (supports listening and turn taking)
- flash cards linked to characters and objects in each story (highlights key vocabulary)
- key vocabulary flash cards to support language games (e.g. Wh cards/opinion cards/sequence cards)
- resource ideas for extended provision (cress seeds & blank wooden spoons)
- RCRI poster & TWiTCH poster (used for display and a visual aid)
- bag and resources organiser (collates resources/cycle)

Practitioners are also provided with:

- the TWiTCH practitioner manual which includes TWiTCH background and theory, details for what each session of the nine book cycles should contain, a one-page summary for each session, planning template to support at-need- children in week 3 and a bank of activity ideas to choose from for week 3 activities
- access to the TWiTCH website for downloadable resources
- access to videos of TWiTCH

Coaches are provided with:

- Coach handbook providing details about the programme, the coach role in the programme and requirements for providing practitioner training and support
- Coach logbook template to capture reflections and conversations during interactions with settings

- Training script and slides for providing training in settings
- Access to resources on the TWiTCH website
- An example setting resource pack
- Training videos

Coach training slides and notes also guide the training and support of coaches.

Tailoring

Implementation - behaviours

An initial meeting between the coach and a senior leadership member before the programme begins is expected to give the setting an overview of how TWiTCH works and what would be expected of the setting and support the coach to learn about the context of the setting and adapt training and support accordingly.

Coaches are encouraged at their training to emphasise the relational nature of the TWiTCH coaching model and to establish a back-and-forth supportive relationship with each of their settings. Delivering training exclusively to members within each setting encourages the establishment of these relationships, so that the setting is more likely to seek help and support and that such support to be tailored to their needs.

The selection of a TWiTCH Champion with responsibility to facilitate and support the delivery of the programme in the setting is expected to encourage greater buy-in and embedding of TWiTCH across the setting.

Regular coach contact and visits to the setting encourage practitioners to reflect on their progress and to set further goals and actions for delivery.

Programme activities through the year encourage continued participation throughout the delivery period.

Network meetings for TWiTCH Champions encourage the establishment of a community of practice across settings, enabling participants to share and solve problems and exchange their own experiences and effective practice.

Implementation – contextual factors

The structured nature of the programme and clearly defined sessions as well as provision of almost all the resources required, facilitates the delivery of the TWiTCH programme and the fidelity of the delivery. During training sessions there is time provided for planning how the programme will be structured and delivered. Coaches understand setting needs and are able to practically support planning at training and through ongoing contact and support. Coaches will check in with settings through email to provide support for any difficulties. Coach visits and observations of practice allow coaches to understand and provide real time feedback and modelling for practitioners. TWiTCH videos allow practitioners to see examples of how TWiTCH can work in practice and to understand the types of child interaction and responses they are trying to facilitate through implementing TWiTCH. These videos can also facilitate TWiTCH delivery when there is staff turnover and new staff who haven't received training need to deliver the programme.

Within the setting, the setting leadership team will enable the delivery of the programme by providing time for practitioners to attend training and coaching sessions. They will also facilitate changes to classroom set-up for the provision of suitable spaces for the small group activities, through considering staffing and ratios to allow the programme to be delivered in small groups. Additionally, the leadership team will support the TWiTCH Champion by providing time to model, and observe TWiTCH practice in the setting. The initial meeting with setting leadership will help senior leadership to understand the expectations and needs of the programme and to facilitate this. The TWiTCH Champion will be vital in supporting the embedding of practice across different practitioners in the setting. Specific training for practitioners and Champion network meetings will help them monitor implementation in their setting and support and improve TWiTCH practice across the setting.

Planned adaptations

Practitioners are expected to incorporate the programme into their normal practice in the way that best suits their setting and children. Settings can choose the timing for delivery of the sessions and who in the setting delivers TWiTCH.

Settings have flexibility in choosing the order of the stories to align best with their curriculum. Exemplar scripts for dialogue are provided but practitioners are expected to adapt these to fit with the context and interests of their children. Suggested vocabulary can be changed by practitioners to that more relevant to their children.

Settings are given flexibility in Week 3 of each cycle to choose activities that fits the needs of their children and the resources available in the setting. Settings are also encouraged to focus and tailor the Week 3 activities to the needs of specific children, especially those that need some additional support. Settings are provided with some guidance in the training to help them select which children might benefit from this support, but the setting has flexibility to decide what this support looks like and how they want to implement it. Coaches are expected to provide feedback and suggestions to practitioners to support their specific needs.

Theory of Change

The TWiTCH (Talk With Tales for Children) programme is designed to address early language gaps that emerge before children enter school and often widen with age. These persistent disparities in children's language development, particularly among those from disadvantaged backgrounds, can limit later educational progress and participation. TWiTCH aims to build practitioners' capacity to support rich language interactions in early years and through this, to improve children's oral language skills before they start school, narrowing the language gap.

The programme is based on the understanding that early years practitioners are key in shaping children's language environments, and it supports practitioners to adopt dialogic storytelling practices and structured small-group language activities that create repeated opportunities for children to practise language, learn vocabulary, express ideas, and reason verbally.

To achieve this, TWiTCH provides a series of key inputs. These include:

- the provision of initial information about the programme and theoretical background through a short video,
- two in-setting training sessions for multiple practitioners delivered by a coach assigned to the setting to explain all aspects of the programme and to help align the programme to the setting's context
- ongoing coaching relationship, visits, observations and meetings to help practitioners embed practices
- full guidance and practical resources for implementing the nine three-week story cycles including a programme manual, story books, structured prompts, a puppet for modelling, a toy to facilitate conversation turn taking.
- Identification of a TWiTH Champion within each setting to support colleagues, model practices, and support with embedding the programme and provision of additional training and online networking meeting support the practitioners in this role.

These inputs are designed not only to develop practitioners' skills but to allow tailoring of the programme to the needs and organisation of the setting and cultivate and support a reflective culture needed for sustained change to practice.

The key programme activities are structured cycles that introduce a new story every three weeks with the activities for each week of the three-week cycle following a similar pattern for each book. In Week 1, dialogic reading encourages children to engage in back-and-forth conversation about different aspects of the story. In Week 2, language-reasoning games provide modelling and space for children to articulate more complex ideas, justify viewpoints, and rehearse newly introduced vocabulary. Week 3 offers targeted consolidation of learning through activities such as story retelling and enhanced continuous provision, ensuring that children who require additional support have tailored opportunities to deepen their understanding.

Through the programme there are a number of ways that these activities are expected to drive change. First, practitioners develop greater confidence, knowledge, and competence in facilitating sustained, high-quality talk with children. Enhanced practitioner capacity is further reinforced through coaching, Champion modelling, and collaborative reflection, creating more consistent implementation and programme-aligned practice across settings. Second, children benefit from repeated exposure to rich vocabulary, structured opportunities to practise reasoning, and engaging story contexts that motivate them to communicate more frequently and more effectively. As children revisit stories, engage with themes, explore characters' choices and are exposed to greater modelling of more complex language, their expressive language and narrative abilities grow, and they begin to use more complex language structures independently.

As these mechanisms take effect, the programme anticipates a series of short-term outcomes. Practitioners report improved understanding of dialogic pedagogy, increased ability to facilitate high-quality interactions, and greater confidence in planning small-group sessions. Children begin to demonstrate greater willingness to participate in group talk, improved vocabulary, better

knowledge of the storybooks and are able to talk about the stories in a deeper way. Children also show emerging improvements in reasoning, explanation, and the ability to use language for thinking, not just for communicating. In the medium term, these shifts contribute to improved oral language outcomes more broadly. As children become more confident and competent in using language to express ideas, negotiate meaning, and understand complex concepts, they enter school with stronger foundational skills. Enhanced language development can also support broader cognitive and socio-emotional outcomes, given the role of verbal reasoning in self-regulation and learning.

The long-term aims of TWiTCH are to reduce language inequalities and improve educational trajectories by ensuring all children—especially those at risk of language disadvantage—experience rich, intentional, and developmentally appropriate language environments. By building practitioner expertise and embedding dialogic practices into everyday routines, the programme seeks to create sustainable change within early years settings that continues beyond the life of the intervention.

The logic model in Figure 1 below summarises these elements and the causal assumptions and contextual assumptions for the programme are set out in appendix B.

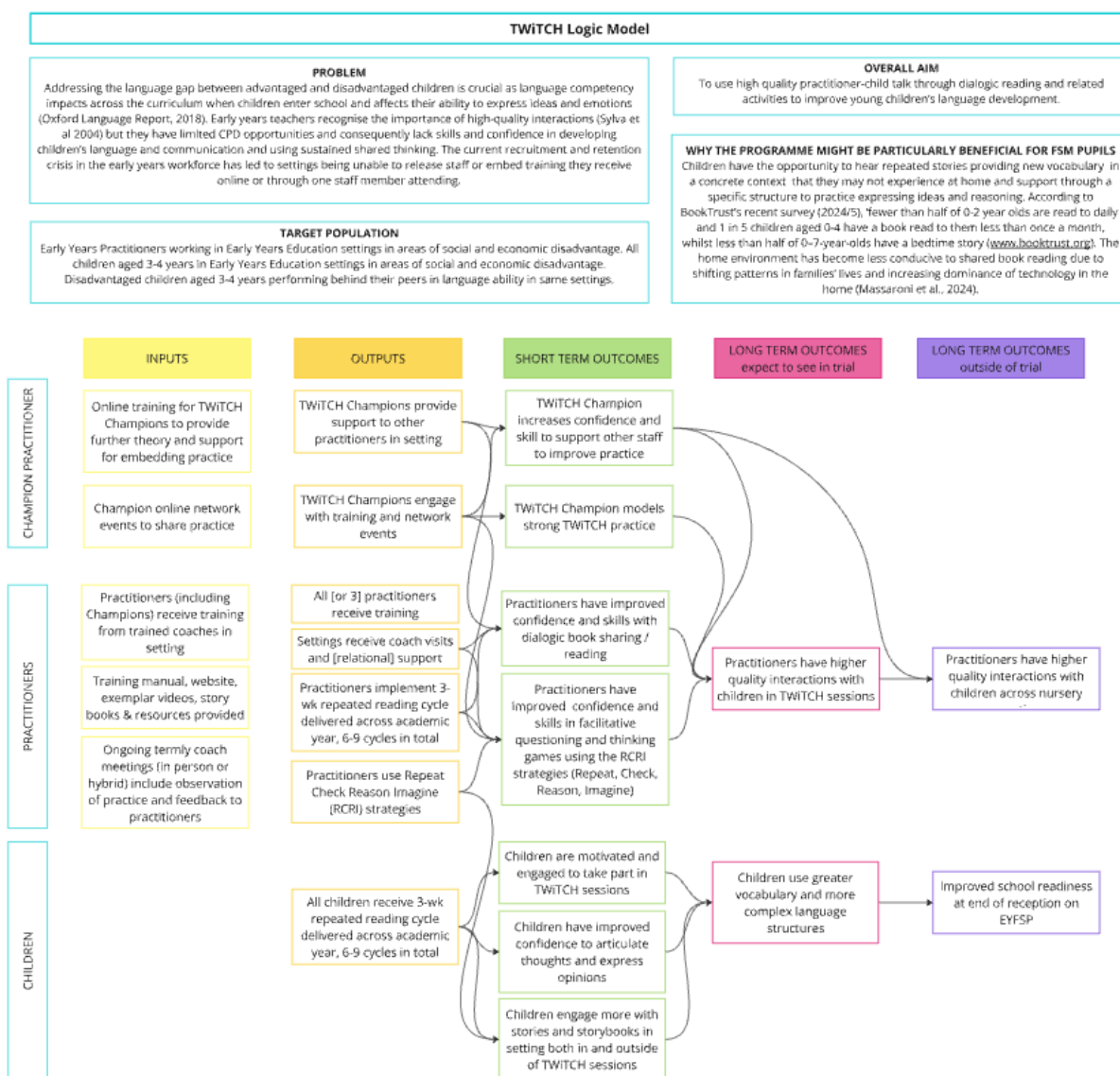


Figure 1. TWITCH Logic Model.

Impact evaluation design

Research questions

Primary Research Question

RQ1. What is the impact of the TWiTCH programme on children's holistic language development measured by the LanguageScreen results of 3–4-year-olds in nurseries receiving the TWiTCH programme in comparison to children in control nurseries receiving business-as-usual?

Secondary Research Questions

RQ2. What is the impact of the TWiTCH programme on children's narrative skills as measured by the Bus Story Assessment of 3–4-year-olds in nurseries receiving the TWiTCH programme in comparison to children in control nurseries receiving business-as-usual?

RQ3. What is the impact of the TWiTCH programme on 1) the subgroup of children eligible for Early Years Pupil Premium, and for 2) the subgroup of children with the lowest language skills at baseline, on their holistic language development and narrative skills comparing children in nurseries receiving the TWiTCH programme to children in control nurseries receiving business-as-usual ?

RQ4. To what extent is the TWiTCH programme more, or less, beneficial depending on children's level of language development at baseline?

RQ5. What is the impact of the TWiTCH programme on nursery practitioners observed story-time practice as measured by coded audio recordings of their story-time behaviour, for nursery practitioners in nurseries receiving the TWiTCH programme in comparison to those in nurseries receiving business-as-usual?

RQ6. What is the impact of the TWiTCH programme on children's Early Years Foundation Stage Profile (EYFSP) assessment at the end of Reception for all children, and those eligible for Early Years Pupil Premium, comparing children in nurseries receiving the TWiTCH programme to children in control nurseries receiving business as usual:

- i) For all EYFSP early learning goals?
- ii) For Communication and Language early learning goals?
- iii) For Literacy early learning goals?

Design

Table 2: Trial design

Trial design, including number of arms		Two-arm cluster randomised controlled trial with random allocation at setting level and business as usual control group
Unit of randomisation		Nursery setting
Stratification variables (if applicable)		Nursery cohort size (<median number of pupils; >=median number of pupils), nursery type (PVI, maintained), geographical region (number dependent on location of recruited settings)
Primary outcome	Variable	Children's holistic language skills
	Measure (instrument, scale, source)	LanguageScreen assessment, latent variable score formed from four subscales (Expressive vocabulary 0-24, Receptive vocabulary 0-23, Listening comprehension 0-16, and sentence repetition 0-14), delivered through app by blinded researchers in setting
Secondary outcome(s)	Variable(s)	Children's oral narrative language
	Measure(s) (instrument, scale, source)	Renfrew Bus Story Test Subscales: 1) Information (i.e., content), 0-52, raw score 2) Sentence length (average number of words in 5 longest sentences) Delivered by blinded researchers in setting
	Variable(s)	School readiness
	Measure(s) (instrument, scale, source)	Early Years Foundation Stage Profile at end of Reception obtained through NPD, 1) 0-17, number of Early Learning Goals (ELG) for which child has met expected standard 2) 0/1, met expected level for all Communication and Language ELGs 3) 0/1, met expected level for all Literacy ELGs
	Variable(s)	Practitioner shared book reading behaviours
	Measure(s) (instrument, scale, source)	Audio recordings of story time practice with small groups coded according to bespoke framework developed in TWITCH pilot evaluation based on Emergent Literacy and Language Early Childhood Checklist for Teachers (ELLECCT) tool. Coding completed by blinded researchers. 1) Frequency per minute of use of all prompt types (open-ended, completion, closed and distancing prompts) 2) Teacher talk percentage during story reading.
	Variable	Children's holistic language skills

Baseline for primary outcome	Measure (instrument, scale, source)	LanguageScreen assessment, latent variable score formed from four subscales (Expressive vocabulary 0-24, Receptive vocabulary 0-23, Listening comprehension 0-16, and sentence repetition 0-14), delivered through app by blinded researchers in setting
Baseline for secondary outcome²	Variable	Children's holistic language skills
	Measure (instrument, scale, source)	LanguageScreen assessment, latent variable score formed from four subscales (Expressive vocabulary 0-24, Receptive vocabulary 0-23, Listening comprehension 0-16, and sentence repetition 0-14), delivered through app by blinded researchers in setting

The design will be a two-armed cluster randomised controlled efficacy trial with randomisation at setting level and a business-as-usual control group. The intervention group settings will receive a thank you payment of £575 for participating in the trial while the control group settings will receive a thank you payment of £1000 in acknowledgement of their participation without receiving the programme. The cluster trial with randomisation at the setting level was chosen due to the design of the TWITCH programme as a whole class programme delivered in groups to all children in the setting which would not allow randomisation at an individual level.

Children's holistic language skills as measured by the LanguageScreen assessment has been chosen as the primary outcome for this trial aligning to the logic model for the programme with the main aim to improve children's language skills. The subscales in this assessment align well with the programme with focus on vocabulary and listening comprehension. The assessment has been standardised and validated with a large UK sample and has been previously used successfully with this age group in other EY language intervention trials. The LanguageScreen assessment is easy to implement consistently and is engaging for children due to being delivered through an app. This assessment is relatively short and showed change in 3–4-year-old children (effect sizes 0.20 for total LanguageScreen score) in an RCT of NELI Preschool, another nursery-based language intervention (West et al, 2024) showing that children improved language skills as a result of the programme. Assessment of children's holistic language skills will be completed at baseline and immediately after programme delivery.

The secondary outcome of children's oral narrative skills, using the Bus Story Assessment, has been chosen due to its specific focus on oral narrative skills expected to improve as part of the dialogic reading aspect of the programme. This assessment goes into greater depth than LanguageScreen on assessing oral language skills and includes aspects of language (complexity of utterance and ability to retell a story) that are more aligned to the TWITCH programme's Theory of Change and may be more likely to show impact. This will be measured only at the end of the programme delivery.

A secondary outcome of children's school readiness as measured through Early Years Foundation Stage Profile data from children at the end of Reception will look at the longer-term impact of the programme on children at the end of their first year of schooling. This will be

² Only for child level secondary outcomes. No baseline used for story time audio recording data

accessed through the National Pupil Database rather than direct data collection from children minimising the burden for settings.

A final practitioner level secondary outcome will be practitioner shared book reading behaviours at the end of the intervention period, measured through coding practitioner audio recordings of story time in all nurseries with a small group of 8-10 children. The coding framework was developed during the TWiTC pilot (Fitzfussell et al., 2025) to observe the frequency of use of specific dialogic tools and was based on the ELECCT tool (Weadman et al, 2022) and TWiTC specific behaviours. The previous TWiTC evaluation showed increased use of TWiTC related behaviours during shared reading from the beginning to the end of the project indicating that practitioners changed their practice as a result of TWiTC. This outcome will be measured in a subsample of 50 settings to minimise burden on practitioners and evaluation cost.

Participant selection

Nurseries

The project will aim to recruit 132 nursery settings to the trial (based on the sample size calculations described below) and a waiting list of approximately 20 settings (an additional 15%) should settings pull out before randomisation or be unable to recruit the required number of children to the project.

Recruitment will focus on state-funded maintained nurseries and PVI settings (aiming for 50% state-funded setting and 50% PVI settings, but recruiting no fewer than 30% PVI settings) across the North West of England, the East Midlands, as well as Yorkshire and the Humber.

To take part, settings must:

- Offer funded places for 3- and 4-year-olds
- Be located within one of the following local authorities:
 - **North West:** Bolton, Bury, Knowsley, Liverpool, Manchester, Oldham, Rochdale, Salford, Stockport, Tameside, Trafford, Wigan, Sefton, St Helens, Widnes, Warrington, Cumbria and Lancashire
 - **East Midlands:** Derbyshire, Nottingham
 - **Yorkshire and the Humber:** Barnsley, Bradford, Calderdale, Doncaster, Kirklees, Leeds, Rotherham, Sheffield, Wakefield, York
- Have sufficient space and staffing to run story time with groups of around ten children
- Have at least 15 children aged 3-4 attending for a minimum of 15 hours per week
- Have a minimum of two staff working with preschool cohorts
- Not be involved in other nursery or reception based EEF projects or EEF Evidence into Action Partnerships during the 2026–27 academic year
- Not previously have taken part in TWiTC training
- Be able to support the evaluation activity requirements.

The delivery team at Sheffield Hallam will lead setting recruitment, with the evaluation team supporting through co-developing materials and attending events.

Recruitment of settings will take a multipronged approach which includes advertising on the EEF's website and through EEF communication, alongside collaborating with several recruitment partners who are in contact with nurseries. These recruitment partners include key personnel within the targeted local authorities, Book Trust and National Literacy Trust charities, and other established relationships and networks of the delivery team. Recruitment partners will be asked to share information about the project in their newsletters and through contacting nursery settings directly. Social media will also be used to advertise the project.

Interested settings are asked to complete an Expression of Interest form and sign up to either a live webinar to find out more about the project, receive a recording of the webinar or to have a one-to-one conversation with the delivery team. They will then be sent the Memorandum of Understanding (MoU) document with full details of the project and asked to complete this and return it if they want to take part in the evaluation project. The delivery team will follow up with contacts who have expressed interest by email and by phone where a MoU has not yet been received. The evaluation team will answer any questions settings might have regarding the evaluation.

Children

The project aims to recruit at least 10 children per participating nursery to the programme with the aim to recruit 1320 children overall. To be eligible to take part children need to:

- be between 3 and 4 years old and due to start school in September 2027,
- attend a participating setting
- attend nursery for at least 15 hours per week
- have parent/carers consent to take part in the evaluation.

Nursery settings will be asked to recruit children to the study using recruitment documentation provided by the evaluation team. Nurseries will be asked to provide information sheets and consent forms to the parents/carers of all children in the 3-4year old room. Parents/carers will be encouraged to complete an online version of the consent form, but where nurseries feel this is not feasible or where online response have been poor, paper copies will be provided, and nurseries will collect paper copies and return these to the research team. The evaluation team will also provide a short video for parents/carers to watch which provides information about the evaluation. Settings will be encouraged to talk to parents/carers about the programme and to answer any questions they have. They will also be asked to translate recruitment materials where appropriate for the families in their settings. Settings who are unable to recruit a minimum of five children for the evaluation would be excluded from the project.

Nursery Practitioners

In each participating nursery, the practitioner expected to be the TWiCH Champion and one additional practitioner who typically leads story time for the 3- and 4-year-old room, will be recruited to complete baseline and endline usual practice surveys and for the practitioner shared reading behaviour outcome measure. The nursery will be asked to nominate these practitioners at the start of the programme so that it is possible to have an equivalent practitioner in both intervention and control settings. Nursery practitioners will be provided with information sheets and consent forms when they complete the initial baseline usual practice survey.

For the practitioner shared reading behaviour outcome measure a subsample of 50 settings (25 intervention and 25 control settings) would be selected, stratified for representativeness across the sample based on socio-economic status (using nursery IDACI ranking data). This will ensure that the sample reflects the diversity of the full trial population and would be statistically valid, representative and feasible, offering better value for money.

Outcome measures

Baseline measures

This study will use the same child language skills assessment (LanguageScreen) at baseline as for the primary outcome which will give the best chance for a strong correlation between baseline and post-test, increasing the power of the trial to detect an effect for the primary outcome. The LanguageScreen assessment will be used with all participating children in September/October 2026 before randomisation. The app-based LanguageScreen assessment will be delivered by independent assessors employed by Navigator Insight working one to one with children in each of the settings. Where a setting has recruited more than five children and a minimum of five children have been assessed, that setting may be randomised before the remaining assessments have been completed. This would allow the scheduling of training to take place for intervention group setting. The remaining baseline assessments would be expected to be completed within the following two weeks and before the intervention training sessions or meetings for the nurseries. The 13th of November would be the cut off point for any baseline assessment to ensure equivalence between baseline assessment in intervention and control group settings. While we acknowledge that allocating settings before all baseline assessments are completed comes with risks which include the potential for limited access to nursery settings for follow-up visits and potential low recruitment rates if settings do not continue to recruit beyond the 5 children, it is necessary to provide time to schedule the training as early as possible to give settings the best possible chance of implementing the full TWiCH programme during the year. The risk will be mitigated by ensuring mop-up visits are booked in with nurseries before randomisation allocation is communicated with settings and that the importance of recruiting more children is continually stressed in communication with settings.

A language skill latent variable score will be created for baseline from the four LanguageScreen subscale scores for use in the analysis. This score will also be used as a baseline covariate in the analysis of the secondary narrative language skill outcome and the Early Years Foundation Stage

Profile outcomes. Further details on LanguageScreen, the justification for using it within the trial and how the assessment is delivered are included in the Primary Outcome section below.

Primary outcome

The primary outcome will be holistic language skills as measured using the LanguageScreen assessment (provided by OxEd Assessment: https://oxedandassessment.com/language_screen). This assessment aligns well with the TWiTH programme theory of change which expects that repeated dialogic reading of story books and improved practitioner dialogic practice would lead to improvement in children's vocabulary and wider language skills. Practitioners reported improvements to communication, vocabulary and wider language skills in the TWiTH pilot study (Fitzfussell et al 2025). LanguageScreen measures four aspects of language skills giving a broad measure of language skills, with three subscales being closely aligned to the intervention: receptive vocabulary, expressive vocabulary and listening comprehension. The LanguageScreen assessment can be delivered with consistency and is engaging for young children due to its app-based delivery. It has also been sensitive to detect changes to language skills with this age group previously in the evaluation of another language intervention (West et al., 2023; West et al., 2024) and has been used in other EEF trials with this age group (Menzies et al., 2024; Menzies et al., 2022).

LanguageScreen is a standardised app-based assessment which will be delivered by an independent assessor visiting the setting blinded to the setting allocation. The LanguageScreen assessment is made up of four subtests: Receptive Vocabulary (23 items where the child chooses which of 4 pictures matches a spoken word (raw score range 0-23)), Expressive Vocabulary (24 items asking the child to name pictures (raw score range 0-24)), Listening Comprehension (listening to 3 stories each followed by a series of questions about the story tapping understanding – 16 items (raw score range 0-16)) and Sentence Repetition (14 items repeating verbatim a series of spoken sentences (raw score range 0-14)). It is administered using an app on a tablet. The app guides the assessment providing full instructions to the assessor and reading aloud all the questions and text for the child. For some items the child taps the screen directly and for others, the assessor decides whether the child has responded correctly or incorrectly and presses the appropriate button. The four assessment subtests are presented in a set order and take around 10-15 minutes to complete. If the tablet is connected to the internet, data from the app is uploaded to the LanguageScreen website automatically. If an internet connection is not available, assessors will be required to sync the data. Results are then generated automatically by LanguageScreen via the portal. A raw score and standardised score for each subtest as well as overall raw and standardised scores are provided. In this project, a latent variable formed from the post-test four subtests raw scores as described in West et al. (2021) will be used for the primary outcome.

LanguageScreen assessment scores correlated strongly ($r=.95$) with a latent variable created from scores on standardised researcher-delivered measures (CELF-Preschool – Expressive Vocabulary subtest and APT information and grammar scores as well as CELF-Preschool Recalling Sentences) in a previous study of more than a thousand participants (West et al., 2021)

showing that the shorter and less intensive LanguageScreen assessment can be validly used to assess language skills.

Navigator Insight will be responsible for the collection of LanguageScreen data both at baseline and post-intervention. They will recruit and train assessors, coordinate the assessment delivery, quality assure assessment delivery and provide the Durham University evaluation team with interim data during the testing windows for checking progress and data quality, and the final assessment data file at the end of baseline and endline testing.

Secondary outcomes

Children's oral narrative language

Children's oral narrative language skills have been selected as a secondary outcome. The TWITCH programme primarily aims to improve oral language (particularly vocabulary and complexity of language structures) through engagement with stories. The Renfrew Bus Story Test (Renfrew, 2010) measures a child's capacity to orally retell a story and provides scores for the type and amount of information included by the child as well as proxies for the complexity of the oral language used. This assessment is more specific to oral language and more in-depth than the LanguageScreen assessment on these elements.

The Renfrew Bus Story Test is a standardised assessment with norms from age three. The assessment is administered one-to-one and typically takes less than 10 minutes. The assessor first reads a scripted story about a bus to the child while showing a corresponding 12-page illustrated picture book. Immediately following the reading, the child is asked to retell the story as accurately as possible while using the pictures as prompts. The retelling is audio recorded to ensure an accurate, verbatim transcription for later scoring. Scoring is based on four measures: i) Information: The number of relevant story facts and details correctly recalled, ii) Sentence Length: The average length of the child's longest five sentences designed as a proxy for complexity of language used, iii) Complexity: the use of independent and subordinate clauses and iv) Independence: the level of cueing from the assessor to get the child to retell the story. This trial will use the i) information and ii) sentence length measures as secondary outcomes as these two scales have been previously shown to be more reliable for scoring across different assessors and on re-retest measures (Education Endowment Foundation, nd) and still align to the aims of the assessment. The information and sentence length scores correlate highly with each other and inter-rater reliability for these subscales is reported as moderately high in the assessment manual. Prankratz et al. (2007) also found that the Bus Story Assessment predicted other language outcomes including oral language scores, expressive and receptive word knowledge, and literacy scores 3 years later.

The Information and Sentence Length subscales will be analysed as separate secondary outcome measures. Although the two subscales are highly correlated, they capture distinct aspects of oral narrative ability: Information reflects recall and organisation of story content, while Sentence Length provides a proxy for expressive language complexity. Analysing these subscales separately allows for clearer interpretation of potential differential effects of the intervention.

The Bus Story Assessment will be delivered at post-test only immediately following the intervention period. It will be delivered by trained assessors employed by Navigator Insight who will visit the settings and assess the children one-to-one. Assessors do not need specialised teaching or speech and language therapist training but do need to have experience working with young children and strong English communication skills. The child's response to the prompts will be audio recorded. The audio recording will then be transcribed and scored by a smaller number of trained assessment moderators at Navigator Insight monitored by an expert in early years assessment.

School Readiness

The intervention logic model indicates school readiness as a long-term outcome of the TWiCH programme with improvements to communication and language expected to lead to greater school readiness and engagement with learning on starting school. The Early Years Foundation Stage Profile is assessed by children's class teachers at the end of a child's first year at school (Reception year) and this covers assessment of 17 Early Learning Goals (ELGs) aligned to the Early Years Foundation Stage curriculum. This data will be accessed through the National Pupil Database rather than directly collected from schools which minimises data collection burden for settings and provides good value for money, with only time required for the statistician and Principal Investigator (PI) to collect, analyse and report on this data.

The EYFSP is an observational measure completed by teachers when children are in the summer term of Reception year (Department for Education, 2025). Teachers rate each child's learning and development against 17 early learning goals (ELGs) across three areas of prime learning (Communication and Language; Personal, Social and Emotional Development; Physical Development) and four specific areas of learning (Literacy; Mathematics; Understanding the World; Expressive Arts and Design) using the following two levels: meeting the level of development expected at the end of the EYFS (expected); or not yet reaching this level (emerging). For any of the ELGs, a score of 'A' may be reported to indicate that a child has not been assessed.

This trial will look at three outcomes based on the EYFSP data. The first will be a global measure of school readiness as measured by the total number of ELGs where the child has met the expected level of development (0-17) across all areas of learning. The other two outcomes will focus on aspects related to the TWiCH programme outcomes: 'Communication and Language', and 'Literacy'. A binary measure will indicate whether the child has reached expected levels of achievement on all ELGs in each of these two areas of Learning. For Communication and Language this includes ELGs on 'Listening, Attention and Understanding' and on 'speaking'. For Literacy, this includes ELGs on 'Comprehension', 'Word reading' and 'Writing'. The child's baseline LanguageScreen latent variable score will be used as a baseline covariate in the analysis.

As this is a longitudinal follow-up with data collected a year after completion of the intervention period, children will have left the nursery to attend school. The evaluation will collect data from nurseries (with parent permission) about where children are expected to attend school and will follow-up directly with parents to collect this information should nurseries not have this

information. This information will be used along with other demographic information collected about children to link to information in the National Pupil Database.

Practitioner shared book reading behaviours

The TWiTCH programme logic model expects that children's outcomes will be positively impacted by the changed practice of nursery practitioners particularly through their use of dialogic reading when sharing storybooks. This is a core component of the programme during the Week 1 activities and repeated cycles of TWiTCH activities is expected to lead to the embedding of these strategies when sharing books with children outside of TWiTCH sessions. These practitioner behaviours will be measured through coding practitioner audio recordings of story time in selected nurseries with a small group of 8-10 children. A subsample of 50 nursery settings (25 intervention and 25 control settings) stratified for representativeness across the sample based on socio-economic status (using nursery IDACI ranking data), will be selected to complete these recordings. In each setting, two recordings will be completed, one by the nominated TWiTCH Champion and one by the nominated other practitioner who completed the evaluation baseline and endline surveys. Practitioners will be asked to share a recording of them sharing a non-TWiTCH picture book of their choice in their normal way, with a small group of children (of up to 10 children) in the class. Practitioners will be advised that the recording should include reading of the book and any discussion of the book that the practitioner has with the children and that recordings are expected to be no longer than 15 minutes although can be shorter depending on the book read. The recording will be coded for practitioner behaviour during the sharing of the story. Coders will undergo a training period where all coders code the same audio recording, meet to discuss their coding and any discrepancies and agree together on the coding for this recording. This is repeated with different recordings until coders have at least 90% agreement on initial coding. Following this training period, 10% of recordings will be double coded with any discrepancies discussed and agreed between the two coders.

The coding framework (see Appendix C) was initially developed during the TWiTCH pilot evaluation (Fitzfussell et al., 2025) and is based on the Emergent Literacy and Language Early Childhood Checklist for Teachers (ELLEECT; Weadman et al., 2022), a validated observational tool designed to document early childhood teachers' extratextual oral language and emergent literacy strategies during shared book reading. The coding framework captures practitioner behaviours across the following categories: Prompts (open-ended prompts, completion and closed prompts, distancing prompts, modelling response to a prompt); Teaching Pedagogy (reviews key ideas, describes story images to students, use of other visual cue, repeats previously presented information, encourages children to reflect on their answers, repeats child's response, provides opportunity to retell the story, provides opportunity to role play story characters, provides corrective feedback, teacher ignores child); Print Instruction (select and stress a word, engages students in print concepts, alphabet knowledge); Phonological Awareness (syllables, rhyming, alliteration, sound awareness); Vocabulary Instruction (provides a definition of vocabulary word, provides an example and non-example of key vocabulary word, engages students in extension activity about key vocabulary word); and Teacher Talk Duration (duration and percentage of teacher talk during story reading and during engagement related to the story). Additional items were included within the framework specifically to capture TWiTCH

pedagogy which do not map directly to the ELLECCT tool specifically within the Teaching Pedagogy section. Coders record the frequency of each behaviour and timestamp occurrences during the recording.

For the impact evaluation, practitioner shared book reading behaviours will be analysed using two pre-specified outcome measures derived from the coding framework, selected on the basis of psychometric evidence and alignment with the TWiTCH Theory of Change.

- Outcome 1: ELLECCT-aligned Prompts section total (frequency per minute). This score sums the frequency of all prompt types coded during the session (open-ended, completion and closed, and distancing prompts) and divides by session duration in minutes to account for variation in recording length. This scoring approach is directly based on the validated ELLECCT Prompts section, which demonstrated excellent intra-rater reliability ($ICC = 0.997$) and excellent inter-rater reliability ($ICC = 0.881$) when applied to shared book reading observations in preschool settings (Weadman et al., 2022). The total prompts score captures the overall level of interactive engagement during story time, how frequently practitioners create opportunities for children to participate verbally rather than passively listen. The TWiTCH programme specifically trains practitioners to increase their use of prompts, particularly open-ended and distancing prompts, during story reading. The directional hypothesis is that practitioners in intervention settings will demonstrate a higher frequency of total prompts per minute compared to control settings, indicating greater interactive engagement consistent with dialogic reading principles. The composition of prompt types (the proportion that are open-ended versus closed) will be reported descriptively in the implementation and process evaluation (IPE) to examine whether TWiTCH shifted the balance toward more cognitively demanding prompt types as intended.
- Outcome 2: Teacher talk percentage during story reading. This is the proportion of time the practitioner spends talking during the story reading portion of the session, expressed as a percentage. This is a continuous, objective measure that requires no categorical judgment about utterance types, only timing of when the practitioner is speaking versus not speaking. It directly operationalises the core dialogic reading principle that the child should increasingly become the storyteller, with the adult's talking role gradually decreasing. In the TWiTCH pilot, teacher talk percentage during story reading decreased from a mean of 82.36% at baseline to 71.21% post-intervention, indicating a meaningful shift toward greater child verbal engagement (Fitzfussell et al., 2025). The directional hypothesis is that practitioners in intervention settings will show a lower percentage of teacher talk during story reading compared to control settings.

Both outcomes will be computed at the practitioner level and then averaged to the setting level for comparison between intervention and control groups using the same framework as the primary analysis.

Wider exploratory analysis of the items in this coding framework and descriptive analysis will form part of the Implementation and Process evaluation. Items on the coding framework that are specific to TWITCH pedagogy and do not map to validated ELLECCT sections (e.g., teaching pedagogy) have not undergone independent psychometric validation and are therefore treated as exploratory descriptive data rather than formal outcome measures and therefore included in the IPE

Delivery of child face-to-face outcomes (LanguageScreen & Bus Story Test) by Navigator Insight

The collection of LanguageScreen and the Bus Story Test data would be carried out by Navigator Insight who will employ a team of early years assessors to conduct assessments one-to-one in nurseries. Assessors will be trained and will deliver assessments blinded to the intervention or control group allocation of the setting. To ensure smooth and effective delivery of these assessments, Navigator Insight will be involved from the beginning of the project and will contribute to the planning and development of the evaluation as regards managing and delivering the child assessments.

The core evaluation team will provide in person training for Navigator Insight key personnel in the delivery and scoring of the assessments. This training will be developed and delivered by Kelly Burgoyne (an expert in child language assessment) and by Vic Menzies, who also has expertise in delivering early years assessment at scale. Following the training, trained Navigator Insight staff will pilot delivery and scoring of the assessments in two settings to ensure full understanding of the data collection tools ahead of the assessor training days and to check adequate time has been planned for assessments. Vic Menzies and/or Kelly Burgoyne would listen to recordings of these assessments to check for quality of delivery and provide feedback to the team in advance of assessor training.

Navigator Insight will recruit at least 30 assessors at each testing timepoint to ensure adequate staffing for delivery to all settings within the testing time period. They will run training for these assessors including online safeguarding training and in-person assessment delivery training in targeted recruitment areas, which will be attended by and monitored by the Durham team with any necessary feedback given at the time. In person training is important to ensure safeguarding and to quality assure each assessor's understanding, manner and delivery of the assessments. Monitoring of assessment delivery will then be done by the Navigator team who check audio recordings of assessors first visit and 10% of assessments beyond this point. Feedback will be given to all assessors after each testing visit and additional training provided, if necessary. All audio recordings will be checked during scoring and inputting.

The scoring of the Bus Story assessment will be completed by specialist Navigator Insight assessment moderators who have been trained by the evaluation team. Assessment moderators have extensive experience of completing assessments themselves and have progressed from this to work as Moderators. To do this, they complete additional training on assessment administration, scoring, quality checking and data processes with Navigator Insight and will also receive direct training from the Core Evaluation Team. Navigator Insight and Durham team members will double score 5% of the Bus Story assessments at the beginning of scoring to ensure

reliability of scoring across scorers. Any discrepancies will be discussed and agreed between all assessors as they come up, and further guidance provided to all assessors to make sure that scoring is completed in the same way. Discussion of discrepancies and double scoring would continue until there is a high rate of reliability of above 80% agreement across all assessors.

In the run up to and during the data collection there will be weekly meetings between the Navigator Insight data collection leads and the evaluation team to plan targets, report progress, flag any issues and monitor completions. Weekly assessment completion targets will be agreed before the start of the assessment period. A live 'Running Totals' spreadsheet will be shared between Navigator Insight and the evaluation team to allow live tracking of assessment progress and reasons for incomplete assessments. Navigator Insight will also share the raw data file once 10% of assessments have been completed and inputted for checking for errors, format and inconsistencies so any errors can be rectified early.

Sample size

Table 3. Sample size calculations

		Overall		EYPP	
		No attrition	With expected attrition	No attrition	With expected attrition
Minimum Detectable Effect Size (MDES)		0.17	0.18	0.31	0.33
Pre-test/ post-test correlations	level 1 (pupil)	0.7		0.7	
	level 2 (class)	-		-	
	level 3 (school)	0.7		0.7	
Intraclass correlations (ICCs)	level 2 (class)	-		-	
	level 3 (school)	0.15		0.15	
Alpha		0.05		0.05	
Power		0.8		0.8	
One-sided or two-sided?		Two-sided		Two-sided	
Average cluster size		10		1.3	
Number of settings	Intervention	66	60	66	60
	Control	66	60	66	60
	Total	132	120	132	120
Number of pupils	Intervention	660	600	86	78
	Control	660	600	86	78
	Total	1320	1200	172	156

Table 4. Expected attrition

	School level	Pupil level
Expected attrition (%)	5	5

The sample size for the TWITCH trial was determined using power calculations (see Table 1), informed by previous EY trials including PACT (Menzies et al., 2024), NELI Preschool (West et al., 2024), and Maths Champions (Robinson-Smith et al., 2024). An intra-cluster correlation (ICC) of 0.15 and a pre/post-test correlation of 0.7 was assumed —values typical for nursery-based interventions targeting language outcomes.

The trial design aims for a minimum detectable effect size (MDES) of 0.18, aligning with meta-analytic findings (Noble et al., 2019), which reported an average effect size of 0.194 for shared reading interventions. This target is conservative and meaningful for the intervention’s goals. The trial design includes 120 settings with 10 children each, yielding an MDES of 0.18. To account for a 10% pupil attrition rate —based on recent EY trials conducted by the project team (Menzies et al., 2024; Robinson-Smith et al. 2024., Brian et al, 2025)—it is planned to recruit 132 settings (an additional 10% settings and pupils), reducing the MDES to 0.17. As the proportion of EYPP children within early-years settings is typically low, the effective cluster size for this subgroup is substantially smaller than for the full sample.

Based on previous EEF EY studies (Bhatti et al., 2024; Brian et al., 2025; Menzies et al., 2022; Menzies et al., 2024; Miller et al., 2023; Robinson-Smith et al. 2018; Robinson-Smith et al., 2019; Robinson-Smith et al., 2024 ; Subosa et al., 2025), an average of 13% of children were eligible for early years pupil premium (EYPP) and therefore approximately 1.3 EYPP children per setting was assumed for this trial. Using the same modelling assumptions as for the main sample—including a two-level cluster design, ICC of 0.15, pre/post-test correlation of 0.7, 80% power, $\alpha = .05$ (two-sided), and equal allocation—an MDES of 0.31 with no attrition was obtained. Incorporating the expected overall 10% attrition, which for EYPP children is likely to translate into loss at both setting and pupil levels, the MDES increases modestly to 0.33.

Baseline assessment would be conducted across the 132 recruited settings and 1320 children with the assumption that between baseline and endline we would be unable to follow-up 10% of children. This may be at setting level (sample sizes assumptions presented in Table 3 below for 10% loss at setting level only) but is likely to be a mix of setting level and individual children level. For budgeting of assessments, we have assumed this loss at 5% setting loss (6 settings) and remainder of loss of children across remaining nurseries bringing the average number of children per nursery at endline to 9.5.

Equal allocation between intervention and control groups (66 settings each) maintains statistical power while ensuring feasibility. Calculations were conducted using PowerUp software, assuming 5% Type I error, 80% power, and a two-sided test.

Further sensitivity calculations (Table 5) explored plausible variations in pre/post-test correlation, cluster size, and number of settings. As expected, higher pre-post-test correlations and larger clusters reduce MDES, improving sensitivity to detect meaningful effects. For

example, with 130 settings and a cluster size of 10, MDES drops to 0.17 when pre/post correlation is 0.7—consistent with the lower end of effect sizes reported in shared reading meta-analyses³ ($d=0.16$ for vocabulary development in Barone et al., 2019). Using the baseline assumptions above considering attrition, where our number of settings is between 120 and 130 and our cluster size is between 7 and 10, but closer to 10 it is likely we can detect an MDES of 0.18 (see scenarios 2a and 2b in Table 5) as planned for. These results support the robustness of the proposed design.

Table 5. Exploring MDES when varying assumptions

Scenario	Pre-post correlation (R^2)	ICC	Cluster size	Minimal Detectable Effect size		
				110 settings	120 settings	130 settings
1a	0.6	0.15	7	0.22	0.21	0.21
1b	0.6	0.15	10	0.21	0.20	0.19
2a	0.7	0.15	7	0.20	0.19	0.18
2b	0.7	0.15	10	0.18	0.18	0.17

Randomisation

The trial statistician, not involved in the day-to-day project activity, will randomise (1:1) nursery settings to either the TWITCH intervention group or the business-as-usual control group after child recruitment and baseline data have been completed. The statistician will only be provided with the setting ID and no identifying features of the settings. Minimisation will be used to ensure balance of sample characteristics across the trial arms (intervention vs business as usual control). Minimisation is a valid approach in randomisation especially in small trials (Altman and Bland, 2005) to ensure only minor differences between groups on variables used in the allocation process. Randomisation will use a dynamic minimisation algorithm with a probabilistic element, following standard practice in EEF-funded cluster trials, with allocation generated using MinimPy, an established Python-based minimisation tool used in previous EEF evaluations such as the EasyPeasy trial. This approach ensures optimal balance across prespecified factors while retaining an element of chance to avoid deterministic allocation, with a 1:1 allocation ratio between the intervention and control groups. The process will be implemented using MinimPy (Saghaei and Saghaei, 2011) alongside Stata (v17 or later) to verify balance, generate the allocation log, and store the random seed for reproducibility.

³ Other meta-analyses have shown larger effects (ranging from effect sizes of 0.26 to 0.62, depending on the outcome, programme type) (Mol et al, 2009; Flack, Field & Horst, 2018; Dowdall et al., 2020).

Randomisation will use minimisation to ensure balance across trial arms on nursery type (PVI; Maintained), cohort size (<median number of pupils; >=median number of pupils) and geographical delivery area/local coach zone as the TWiCH programme delivery is expected to be impacted by these factors. Nursery type is used as a minimisation factor to ensure the type of early years providers are well balanced between the intervention and control groups before the intervention; (i.e., staff qualifications and children's disadvantaged status are often different between PVI and maintained/school nursery settings). The nature of these different settings often leads to different organisation of the setting which may impact on how and when the TWiCH programme can be implemented. The TWiCH pilot evaluation indicated that larger settings could find implementing TWiCH with their full cohort more difficult, so minimisation based on cohort size is used to ensure larger settings are balanced across the intervention and control groups. All recruited settings will be assigned to a delivery area prior to randomisation. Delivery areas will be defined in consultation with the delivery team, based on coach location, travel feasibility and setting recruitment geography. This will ensure that intervention and control settings are balanced within geographical areas, while supporting the feasible distribution of intervention settings across areas where coaches are available. Coaches will be assigned to intervention settings after randomisation by the delivery team, taking into account coach capacity and travel constraints. Coach allocation will not determine trial allocation but will be recorded for implementation monitoring and process evaluation."

Nurseries will be randomised on a weekly rolling basis after child recruitment and baseline data collection is complete for the minimum number of children in each setting. A dataset containing all nurseries and the populated minimisation variables will then be prepared, after which the MinimPy algorithm will be run using a fixed seed to ensure reproducibility. This process will generate the allocation list assigning nurseries to either the intervention or control group. Mop up assessment visits may be done after randomisation but before any intervention training begins (expected to be no more than 20%) to maximise the time for the intervention delivery. All nurseries will be informed of their random allocation via a letter emailed to the nursery contact from the delivery team. A weekly email to the delivery team will provide the details of which settings have been randomised to which group.

Statistical analysis

The statistical analysis will follow the most recent EEF Statistical Analysis Guidance (EEF, 2022). All analyses will be conducted on an intention-to-treat (ITT) basis, using two-sided significance at the 5% level. A detailed data analysis strategy will be included in the SAP. Below we provide a summary of the planned statistical analysis of primary and secondary outcomes, subgroup analysis, analysis in the presence of non-compliance and additional analyses.

Baseline data will be summarised by trial arm and presented descriptively. Effect sizes based on the difference between the intervention and business as usual groups at the post-test will be presented as Hedges' *g* with 95% confidence intervals. The ICC of the pre-test and post-test for the primary outcome will also be presented.

Primary analysis

The primary analysis will investigate any difference in the primary outcome (children's holistic language skills as measured by the latent variable score created from the four LanguageScreen subscales at the end of the intervention period). The analysis will take the form of a linear mixed model at the child-level with primary outcome latent variable score as the dependent variable. Group allocation, baseline score and minimisation factors will be included as fixed effects in the model. Adjustments for clustering at the setting level will be made by including the setting as a random effect.

The latent variable score will be constructed using weightings extracted from the factor loadings matrix of a confirmatory factor analysis (CFA) conducted on the post-test data. These loadings will be treated as fixed and known for all subsequent analyses, and the latent language score will be calculated as the weighted sum of the four subscale scores. Model fit for the CFA will be evaluated using established thresholds: root mean square error of approximation (RMSEA) < 0.08, standardized root mean square residual (sRMR) < 0.08, comparative fit index (CFI) $\geq$ 0.90, and Tucker–Lewis index (TLI) $\geq$ 0.95 (Hu & Bentler, 1999; Kline, 1998). To ensure consistency in measurement structure over time, the same weightings derived from the post-test CFA will be applied to baseline scores; a separate CFA will also be conducted at baseline to check whether the factor structure and loadings are consistent with those observed at post-test.

Secondary analysis

Secondary analyses will investigate any differences in secondary outcomes: Children's oral narrative skills (measured using the Bus Story Assessment), Children's school readiness (measured using Early Years Foundation Stage profile data at the end of Reception) and frequency of practitioner's shared book reading behaviours measured through the coded audio recordings of practice. Appropriate techniques will be used for the type of data collected (e.g., continuous or categorical data).

Children's oral narrative skills

The raw scores of the i) information and ii) sentence length Bus Story Test subscales will be used to represent children's oral narrative skills attainment at post-intervention. Each subscale will be analysed separately which offers the advantages of greater interpretability, and methodological transparency appropriate for secondary outcomes in an RCT. This secondary outcome analysis for both subscales will take the form of a linear mixed model following the same specifications used in the primary analysis comparing the intervention and control group performance on each subscale at post-intervention. LanguageScreen latent variable scores at baseline will be used as a covariate in the model.

Children's school readiness at the end of reception

A number of different outcomes will be created for the school readiness outcome measure. The number of early learning goals that were rated as at expected will be analysed as measure of overall school readiness (0-17). This secondary outcome analysis will take the form of a linear mixed model following the same specifications used in the primary analysis comparing the

intervention and control group performance on each subscale at post-intervention. LanguageScreen latent variable scores at baseline will be used as covariates in the model.

Analysis will also look specifically at the impact of the TWITCH programme on whether the three Communication and Language ELGs were rated as expected (binary all met or not all met), and separately at whether the three ELGs for Literacy were all rated as expected (binary: all met or not met). As both these outcomes use a binary outcome variable, analysis will take the form of a mixed-effects logistic regression model with group allocation, baseline score and minimisation factors will be included as fixed effects in the model. Adjustments for clustering at the setting level will be made by including the setting as a random effect.

Practitioners shared book reading behaviours

The i) frequency of use per minute of all coded prompts by practitioners and the ii) percentage of teacher talk during the shared book reading will be used as separate dependent variables and will be analysed at the practitioner level using a linear mixed model comparing control group and intervention group outcomes. Group allocation, and minimisation factors will be included as fixed effects in the model. Adjustments for clustering at the setting level will be made by including the setting as a random effect. There will be no baseline data for this outcome.

Estimation of effect sizes

Effect sizes will be calculated using Hedges' g , following the approach used in EEF cluster-randomised trials. The adjusted mean difference between groups, estimated from the multilevel model, will be divided by the square root of the total variance (including both within-cluster and between-cluster components) obtained from an unconditional multilevel model, consistent with the method described by Hedges & Hedberg (2007). Confidence intervals for the effect size will be calculated using the Hedges (2007) formula.

Sub-group analyses

Subgroup analysis will be conducted for:

- the Early Years Pupil Premium (EYPP) eligible group with data collected from nurseries at the end of the intervention period
- children with the lowest language scores at baseline⁴

Subgroup analyses will be done via the inclusion of the status and an interaction term between status and allocation in the primary analysis model, and by repeating the primary analysis in the specified subgroups of children considering the similar set of covariates as used in the primary outcome analysis model. Subgroup analysis will only be completed for the primary outcome.

⁴ This subgroup will be discussed and defined in the SAP.

Analysis in the presence of non-compliance

A setting level compliance descriptor has been developed with the delivery team defining compliance across all elements of the TWiATCH programme (Table 6). Within this descriptor, five core elements expected to be the causal mechanisms have been identified for use in the compliance analysis. These are:

- At least two staff members attend both in-setting training sessions (scored 0 or 1)
- Setting delivers at least 7 cycles of TWiATCH (scored 0 or 3⁵)
- TWiATCH is delivered to groups of 10 children or fewer (scored 0 or 1)
- Three coaching sessions take place during intervention delivery (scored 0 or 1)
- Specified TWiATCH activities delivered over three weeks of cycle (scored 0 or 1)

Each nursery will be given a compliance score between 0 and 7 for their level of compliance with the programme. Data will be collected from coach logbooks, training attendance logs and the post-intervention practitioner survey. The other aspects of programme compliance will be explored in the Implementation and Process evaluation exploring the extent to which each aspect of the programme is delivered and the reasons for variations in compliance.

Table 6. TWiATCH Compliance Descriptor

TWiATCH Programme Component	Data collection
Core compliance measure	
At least two staff members attend both training sessions	Training attendance logs
Setting delivers at least seven cycles of TWiATCH (triple weighting)	DU post-intervention survey
TWiATCH is delivered to groups of 10 or fewer	Coach logbook DU – post intervention survey
Three coaching sessions take place during the intervention delivery	Coach logbook
Specified weekly TWiATCH activities (dialogic reading, choice-based language games and continuous provision activities) delivered over three weeks of cycle	Coach logbook DU-post intervention survey
Facilitative components (explored in IPE)	

⁵ Higher weighting due to importance of increased dosage for expected impact

Engages with online training video	Coach asks at training – coach logbook Check video analytics
Coach and senior leader meet before programme	Coach logbook
TWiTCH Champion attends Champion training	Attendance logs from SHU
Cycle 1 delivered between 2 training sessions	Coach logbook
Uses repeated reading of book at least 4 days during week 1 of cycle	DU – post intervention survey
Implements 3 choice-based language sessions during week 2	DU – post intervention survey
Does at least 1 planned continuous provision activity for children during week 3	DU-post intervention survey
Some targeting of individuals/small groups in week 3	DU – post intervention survey
Delivers TWiTCH to whole cohort/room	Coach logbook
Champions engage with at least 1 (of 2) network meetings	Attendance records
Use of RCRI strategies reported by coach	Coach logbook (from observation)

Complier Average Causal Effect (CACE) analysis would explore for the primary outcome how the impact of the evaluation varies depending on the level of compliance with the programme (compliance score). The compliance score will be used as a continuous variable.

CACE analyses will use a Two Stage Least Square (2SLS) approach with group allocation as the instrumental variable for the compliance indicator. Results for the first stage (which predicts the compliance indicator using treatment allocation as an instrumental variable alongside all other covariates included in the second stage) will be reported alongside i) the correlation between the instrument and the endogenous variable; and ii) an F test (EEF, 2022).

Separate to the above analysis, to investigate whether the programme is more effective for children who attend the nursery for more hours and are therefore more frequently exposed to intervention activities, analysis will explore how the programme impact differs for i) those who attend nursery for at least 30 hours per week, ii) those that attend between 15 and 20 hours per week and iii) those that attend between 20 and 30 hours per week. Planned nursery attendance collected at project sign up will be used as a proxy measure for attendance for this analysis. These analyses will be conducted by including an interaction term between group allocation and attendance category in the primary multilevel model. If the interaction is statistically significant, stratified estimates will be presented for each attendance band. As these analyses are exploratory and based on proxy measures of exposure, effect estimates will be interpreted with appropriate caution.

The setting level and child level compliance will be analysed separately to aid interpretation of the factors which have the greatest influence on outcomes.

Additional analyses and robustness checks

Sensitivity analysis would explore the fit of the latent variable model for the LanguageScreen primary outcome through structural equation modelling (SEM). SEM will be specified in which the four LanguageScreen subscales load onto a single latent holistic language factor, with treatment allocation modelled as a direct effect on the latent outcome. Results from the primary linear mixed model using the pre-computed weighted latent score will be compared with estimates obtained from the SEM analysis. Consistency in the estimated intervention effect (in terms of direction and magnitude) across the two approaches will be used to assess the robustness of conclusions to the analytic method used for the primary outcome.

Any differences in effects of the intervention for PVI compared to maintained nurseries across settings will be explored for all primary and secondary outcomes.

As raw scores will be used for analysis, we will conduct sensitivity analysis of age adjustment to check whether the inclusion of age variable into the primary outcome analysis model makes a notable difference in the effect size estimates.

Exploratory analysis will look in more depth at the impact of the programme on the specific ELGs across the EYFSP and how they align with the logic model. This will provide more nuanced information about the specific impact of the programme and the causal mechanisms.

Missing data analysis

The amount of missing primary baseline and post-test outcome data will be summarised for both arms, and reasons for missing data and patterns of missing data will be explored and provided. The impact of missing data on the primary analysis will be assessed by repeating the analysis on a data set where missing data has been completed using multiple imputations should the missing data be above a threshold set out in the SAP. Missing data in the baseline and outcome measures will be accounted for using full information maximum likelihood estimation techniques (Cham et al. 2017).

Estimation of ICC

The intra-cluster correlation coefficient (ICC) associated with the nursery school for the outcomes (both pre- and post-test primary outcome where available) will be presented. The ICC at post-test will be computed for the empty linear mixed model with nursery settings as a random effect (i.e. one without covariates). The ICC at pre-test will be calculated using a linear mixed model with pre-test as the outcome and setting as a random effect.

Implementation and process evaluation (IPE) design

The IPE will focus on four dimensions: implementation and fidelity, usual practice, and causal mechanisms and assumptions for the programme and scale up. Within these areas specific targeting of aspects raised in the pilot evaluation (Fitzfussell et al, 2025) will be explored which include:

- consistency of delivery of training and support
- size of group for delivering TWiCH
- dosage of the intervention through child attendance and whether this impacts on outcomes (complementing the impact evaluation)
- how nurseries organise the day and their space to facilitate delivery to all pupils
- unintended consequences including practices that were stopped to deliver TWiCH.

The IPE will complement the impact evaluation by supporting explanation of the observed effects of the TWiCH intervention. Qualitative and quantitative data from relevant stakeholders in the IPE will be triangulated with the impact evaluation outcome data and explored within the context of the ToC. Exploration of fidelity and compliance through the IPE will feed into the interpretation of compliance in the CACE analysis.

Research questions

Implementation and fidelity (IF)

IPE-RQ1: How consistently is TWiCH training and support delivered across settings and coaches?

IPE-RQ2: How closely is TWiCH implemented as planned across the 9 cycles and week cycles? What explains variation in fidelity? How is the programme adapted?

IPE-RQ3: Do all children receive the full TWiCH content, and how does variation affect engagement and outcomes?

- How do 'non-full time' children access missed content if they do? Do 'catch-up' strategies vary by setting type (PVI vs school) or by practitioner experience or expertise?

IPE-RQ4: What factors influence disadvantaged children's access to and engagement with TWiCH?

IPE-RQ5: What barriers do settings face in delivering TWiCH?

IPE-RQ6: What supports effective programme delivery?

Usual Practice (UP)

IPE-RQ7: How does TWiTC differ from usual nursery practice in structure, content, and grouping?

IPE-RQ8: What do nurseries stop doing to deliver TWiTC?

IPE-RQ9: What is the control group's usual practice during the intervention period?

IPE-RQ10: How are children (including disadvantaged children) with poor language skills typically supported, and how does TWiTC change this?

Causal Mechanisms and Contextual Assumptions (CMCA)

IPE-RQ11: Is TWiTC perceived as suitable for all 3–4-year-olds?

IPE-RQ12: How does group size affect implementation, engagement, and outcomes?

IPE-RQ13: Who receives extra support in week 3, and what impact does TWiTC have on them?

IPE-RQ14: How effective and appropriate is the TWiTC training and support model?

IPE-RQ15: How do nurseries organise time and space to deliver TWiTC?

IPE-RQ16: How does TWiTC influence practitioner and nursery practice beyond the programme?

IPE-RQ17: What impacts—expected or unexpected—do stakeholders perceive from TWiTC? Does TWiTC impact on relationships between practitioners and children and between children in the class? Does TWiTC impact on children's wider relational skills and language?

IPE-RQ18: How do TWiTC Champions support other practitioners?

IPE-RQ19: How does TWiTC change nursery practitioners' confidence, knowledge and practice regarding dialogic book reading and supporting children's language skills?

Scale up (SU)

IPE-RQ20: What changes have been made to TWiTC to allow it to be scaled up for this efficacy trial and what future changes would be needed to allow wider scale up to a larger number of schools beyond the efficacy trial?

Research methods

The IPE uses a mixed methods design with data collection at multiple time points. It includes both pre and post-intervention survey data collection from across the full sample and in-depth interviews and focus groups from a selected sample of participants, observations of TWiTC programme training and delivery and case studies towards the end of the programme. Details of the IPE methods are set out in Table 7.

Table 7. Implementation and Process Evaluation Research Methods

Administrative Data					
<i>Research Methods</i>	<i>Who?</i>	<i>What will this tell us</i>	<i>IPE area (RQ)</i>	<i>When</i>	<i>Analysis</i>
Training attendance and Coach logbook delivery records from coaching visits to setting	Data about all settings from DT. Logbook prompts will be developed with DT for coaches to use to ensure good quality data.	Enable monitoring of fidelity and compliance to the programme across all settings with minimal data collection burden.	IF (RQ1, RQ2, RQ3, RQ5)	Throughout delivery.	Compiled to create setting compliance scores. Descriptive statistics provided for implementation of core elements across settings. Qualitative analysis of log books for barriers, facilitators, adaptations and coach support.
Data available about nurseries from online sources	Data for all nurseries	Provide contextual information about nurseries to inform interpretation of results. Support recruitment.	CMCA, UP (RQ7, RQ14, RQ15, RQ16, RQ19)	At sign up.	Use in description of participating settings, interpretation and to create potential subgroups for exploratory analysis e.g. by type of setting, setting characteristics, rural/urban, Ofsted rating.
Audio recording coded data					
<i>Research Methods</i>	<i>Who?</i>	<i>What will this tell us</i>	<i>IPE area</i>	<i>When</i>	<i>Analysis</i>
Coded audio recordings of practitioner behaviour during story-time practice using bespoke coding framework (Fitzfussell et al., 2025) (collected for IE)	Nominated TWiTH Champion and practitioner in 50 settings (25 control group and 25 intervention group) stratified by IDACI scores.	To look at whether specific practitioner behaviours aligned to the TWiTH programme theory of change are more evident in the intervention group compared to	CMCA (RQ16)	Endline only	Descriptive and exploratory statistical analysis comparing practice in intervention group with control group.

		the control group and are embedded in TWiCH practitioners wider practice.			
Surveys					
<i>Research Methods</i>	<i>Who?</i>	<i>What will this tell us</i>	<i>IPE area</i>	<i>When</i>	<i>Analysis</i>
Online practitioner usual practice surveys incorporating confidence and knowledge scales.⁶	Expected TWiCH Champion and one practitioner of 3–4-year-olds in all settings	Provide data on practitioner characteristics , usual nursery classroom and CPD practice and other training undertaken. At endline they will also capture views on programme delivery, required resources and impact for intervention settings and usual practice during the trial in control settings.	IF,UP, CMCA (RQ1, RQ2, RQ3, RQ5, RQ7, RQ8, RQ9, RQ10, RQ11, RQ13, RQ14, RQ15, RQ16, RQ17, RQ18, RQ19	Baseline, Endline	Descriptive statistics exploring by group allocation, setting type and project role. Inferential statistics for confidence, knowledge and practice scales comparing intervention group & control group. Deductive/inductive coding for qualitative responses.
Online senior leader surveys	A senior leader in all settings	Provide nursery level usual practice and contextual information, motivation for signing up to the programme, resourcing for the programme, impact on the setting (at endline),	UP, CMCA. (RQ6, RQ7, RQ8, RQ9, RQ10, RQ11, RQ14, RQ15, RQ16, RQ17, RQ18, RQ19)	Baseline, Endline	Descriptive statistics exploring by group allocation, setting type. Deductive/inductive coding for qualitative responses.

⁶ Three scales of confidence, language supporting practice and story-time practice for inclusion in surveys already developed as part of the TWiCH pilot evaluation.

		further embedding into nursery practice.			
Short post-training survey	All nursery practitioners who attended training	Views of training, knowledge and confidence gains for programme delivery, resourcing required to attend training	CMCA (RQ14)	After training sessions.	Descriptive statistics exploring by setting type and project role. Deductive/inductive coding for qualitative responses.
Observations					
<i>Research Methods</i>	<i>Who?</i>	<i>What will this tell us</i>	<i>IPE area</i>	<i>When</i>	<i>Analysis</i>
Semi-structured observation of coach training	3 x 'train the trainer' coach training sessions	Expected delivery practice for coaches	IF (RQ1, RQ2)	Before programme delivery	Deductive analysis
Semi-structured observation of coach-led training	6 observations (3 for each session of training) of randomly selected coaches across a variety of setting types and characteristics	Consistency of training delivery and fidelity to the programme. Also, will capture engagement with training.	IF (RQ1)	Early programme delivery.	Deductive analysis to look at consistency of training, level of engagement
Semi-structured observation of mentoring sessions	6 observations of randomly selected coaches	Consistency of delivery and fidelity to the programme.	IF (RQ1, RQ2, RQ3,	Throughout delivery including 2 observation	Deductive analysis to look at consistency of training, level of

including observation of TWiATCH lesson delivery	across a variety of setting types and characteristics .		RQ4, RQ5, RQ6, RQ12, RQ15)	s of each of the three mentoring sessions	engagement, fidelity to programme.
• Semi-structured observation of TWiATCH champion training and online network events	Up to 4 observations.	Role of these sessions within the programme and engagement with them. Challenges and facilitators to delivery across settings.	IF, CMCA (RQ1, RQ2, RQ5, RQ6, RQ14)	Throughout delivery.	Deductive analysis looking at engagement with session, challenges and facilitators to delivery.
• Semi-structured observation of TWiATCH practice (Case study settings)	2 x best practice high fidelity case study settings 2 x low fidelity case study settings	To assess what the intervention looks like, how it works, how closely it aligns with programme materials and the engagement of children in the session.	IF, CMCA, (RQ2, RQ3, RQ4, RQ5, RQ6, RQ12, RQ15)	Towards end of programme delivery.	Deductive analysis of programme delivery, fidelity and engagement.
Interviews/Focus groups					
<i>Research Methods</i>	<i>Who?</i>	<i>What will this tell us</i>	<i>IPE area</i>	<i>When</i>	<i>Analysis</i>
• Semi-structured interviews (two time points) phone/online	TWiATCH Champion in 8 settings chosen to represent a variety of setting characteristics .	To understand practitioners' views of the programme, including training and acceptability of materials for their children. To find out about the organisation of TWiATCH across settings. To understand	IF, CMCA, UP (RQ2, RQ3, RQ4, RQ5, RQ6, RQ7, RQ8, RQ10, RQ11, RQ12, RQ13, RQ14, RQ15, RQ16,	Early in programme delivery, towards end of programme delivery	Deductive/inductive coding (Bingham, 2023)

		barriers, facilitators and adaptations made. To look at perceived impact of the programme and how TWITCH fits with usual practice.	RQ17, RQ18, RQ19)		
Semi-structured interviews in person (Case study settings)	Champion, practitioner & senior leader in 2 x best practice high fidelity case study settings 2 x low fidelity case study settings	To explore impact, barriers and facilitators, and necessary conditions for success.	IF, CMCA, UP (RQ2, RQ3, RQ4, RQ5, RQ6, RQ7, RQ8, RQ10, RQ11, RQ12, RQ13, RQ14, RQ15, RQ16, RQ17, RQ18, RQ19)	Towards end of programme delivery	Deductive/inductive coding (Bingham, 2023)
Semi-structured interviews (two time points)	Developer team	To explore any adaptations made, perceptions of programme delivery, perceived impact, future delivery model.	IF, CMCA, UP, SU (RQ1, RQ2, RQ5, RQ6, RQ7, RQ10, RQ12, RQ14, RQ17, RQ18, RQ20)	Early in programme delivery. After programme finishes.	Deductive/inductive coding (Bingham, 2023)
Focus group (two time points) online	Focus groups at two time points with all coaches	To explore views and experiences of programme delivery and	IF, CMCA (RQ1, RQ2, RQ3,	Early in delivery following training delivery,	Deductive/inductive coding (Bingham, 2023)

		reflections of programme PD model.	RQ4, RQ5, RQ6, RQ12, RQ13, RQ14, RQ15, RQ17, RQ18, RQ19, RQ20)	one towards the end of delivery	
Cost and scale up workshop					
<i>Research Methods</i>	<i>Who?</i>	<i>What will this tell us</i>	<i>IPE area</i>	<i>When</i>	<i>Analysis</i>
Cost and scale up workshop	Core DT and ET	Establish programme costs, 3-year delivery model, model for scale up and assumptions for cost calculations	Cost evaluation (RQ20)	After programme delivery.	In line with EEF cost evaluation guidance.

Survey measures

Baseline and Post-intervention Practitioner surveys

All participating nurseries will be asked to nominate a potential TWiTCH Champion before being randomised to intervention or control group. This nominated Champion along with one other nursery practitioner in each setting who would be expected to deliver TWiTCH, will be asked to complete baseline surveys, before randomisation is completed (baseline), and after the intervention delivery period (post-intervention) (the sample will include practitioners in both intervention and control group settings). Surveys will be hosted online using Qualtrics and the links provided by email to practitioners.

These surveys will provide data on practitioner characteristics, usual nursery classroom and CPD practice and other training undertaken at baseline (and endline when appropriate). At endline surveys will also capture areas of programme delivery, required resources and perceived impact for intervention settings, and usual practice during the trial in control settings. Surveys will contain quantitative and qualitative questions in relation to these topics.

At both time points the surveys will also include short self-report measures of 1) practitioners' confidence in supporting children's language skills, 2) practitioners' usual practice when interacting with children in the nursery, and 3) practitioners' usual practice during storytelling sessions. These measures are included as the TWiTCH programme theory of change expects practitioner skill and knowledge in supporting children's language development to improve as

part of the programme. These measures were developed and used previously in the TWiTC Pilot evaluation (Fitzfussell et al., 2025) and were adapted from the Practitioner Confidence and Skills in Supporting Children's Language Measure (Lindorff, Sylva, Ereky-Stevens & Joseph, 2022) tailoring it towards expected skills and practices aligned to the TWiTC programme. For items in each scale, participants respond using a five-point Likert rating to rate level of confidence for scale one and frequency of activity for scales two and three. Scale reliability calculations in Fitzfussell et al. (2025) show scales had high internal reliability of 0.88 or above. The pilot evaluation found positive significant differences between scale scores before and after completing the TWiTC. This study will allow comparison of scores between intervention and control groups allowing stronger causal claims to be made.

Baseline and Post-intervention Senior Leader surveys

A senior leader in each participating nursery setting will be asked to complete a short online survey hosted on Qualtrics, at baseline and post-intervention. This survey is designed to collect nursery level contextual information including motivations for signing up to the programme, setting organisation and staffing, information about usual practice and training. At post-intervention, surveys for intervention group settings will also include questions around resourcing for the programme, impact on the setting and further embedding of the programme into nursery practice. Where the senior leader has also been delivering TWiTC, relevant questions from the practitioner survey will also be included so that the senior leader only needs to complete one post-intervention survey.

Post-training surveys

All intervention group practitioners who attend the in-person training sessions will be asked to complete a short online survey, hosted on Qualtrics, at the end of each session with time provided for completion. These surveys aim to capture knowledge gains regarding TWiTC strategies, practitioner confidence to deliver the programme, acceptability/appropriateness of the training, and any feedback practitioners had about the training sessions. While the surveys will be developed by the evaluation team, they may include specific questions provided by the delivery team to support them with gathering feedback about the training. This survey will be cross sectional, using mostly quantitative Likert scale responses but providing the space for practitioners to elaborate where necessary. This information will help the delivery team to understand and adapt their training moving forwards.

Interviews

All interviews will be audio recorded with permission from the participants and transcripts made for data analysis.

Remote Practitioner interviews (8 settings)

Semi-structured phone/online interviews will be undertaken with the nominated TWiTC Champion in eight purposively selected settings. Settings will be chosen to represent a variety of setting characteristics with consideration of type of setting, location, setting cohort size, Ofsted rating and level of deprivation. Interviews will be undertaken at two time points (after delivery of

two to three TWiTCH cycles and towards the end of the programme) with the same Champion adding a longitudinal component to the design allowing investigation of any changes to delivery as time goes on and as children develop. These interviews will be designed to understand:

- practitioners' views of the programme, including training and acceptability of materials for their children.
- how TWiTCH is organised and delivered across settings and the adaptations being made
- barriers and facilitators to implementing TWiTCH and seeing impact
- the perceived impact of the programme
- how TWiTCH fits with usual practice.

Interviews will take around 20 minutes each.

Case study practitioner interviews (4 settings)

Semi-structured interviews will take place with setting staff involved in the delivery of TWiTCH in selected case study settings towards the end of programme delivery (these will be different settings to those involved in the remote practitioner interviews). Two case study settings will be selected as best practice settings where the programme is being delivered with high fidelity, and two case study settings will be selected where the programme is not being delivered with high fidelity based on implementation fidelity data collected through the coach logbooks. In case study settings we will invite the TWiTCH Champion to take part in a 15–20-minute interview during an in person visit to the setting which will also include observation of TWiTCH practice. Where settings are willing and able to facilitate it, we will also invite another practitioner and a senior leader to take part in an interview. It is expected that it may be more difficult for settings with low fidelity to release staff to take part in these interviews and by only requiring the TWiTCH Champion to participate this may make settings more willing to engage with the IPE.

These interviews will explore programme impact, delivery, barriers and facilitators and the necessary conditions for success of TWiTCH.

Delivery team interviews

Semi-structured interviews will take place with members of the core delivery team at two time points during the project (early into programme delivery and after the programme delivery is complete). The first interview will explore delivery of coach training and setting training, any adaptations made, any challenges faced, and how it compared to what was expected. The second interview will explore reflections on delivery across the whole project, changes that would be made for future delivery, perceived impact and will consider how TWiTCH could be delivered in the future. These interviews are expected to take around one hour at each time point with one interview taking place in person and the other being conducted remotely.

Focus groups will also be undertaken with all TWiTCH coaches at two time points during the project (early in programme delivery after providing training to settings and towards the end of the

programme delivery). These focus groups will explore the views and experiences of the coaches regarding the training and delivery of TWiTCH across the settings they support and their reflections on the programme PD model. Focus groups are expected to take between 45 minutes and one hour.

Observations

Coach training sessions

Semi-structured observations of the three coach training sessions will be undertaken to understand the expected delivery practice for coaches across the programme which will inform judgements of implementation fidelity when observing setting training. Observations will be structured to look at training content, how closely the training follows the pre-planned content, the engagement of the participants, questions raised by participants and overall reflections of facilitators or barriers to coach learning or engagement with the training.

Coach led setting training

Six semi-structured observations of setting training (led by coaches) will be undertaken (three observations for each session of training) for six randomly selected coaches across a variety of setting types and characteristics. These will be undertaken in settings not participating in remote practitioner interviews to minimise setting burden. Observations will be conducted using an observation template to capture training content, how closely the training follows the pre-planned content, nursery context and how training was adapted to suit context, engagement of participants with training and reflections of observer about fidelity to the programme. Across training observations, we will look at the consistency of training delivery across coaches.

Mentoring sessions

Six semi-structured observations of mentoring sessions including an observation of TWiTCH lesson delivery (where applicable) will be undertaken across randomly selected coaches in a variety of setting types and characteristics. These will be settings where observation and interviews have not been undertaken. Two observations of each of the three mentoring sessions through the programme will be observed. During each observation an observation template will be used which will capture:

- the structure of the mentoring session and content and its alignment to training and coach handbook
- fidelity of the TWiTCH lesson delivery to the programme during the session observed, engagement of children in the session, the input of the coach in the session
- feedback from coach to practitioner
- questions/discussion points raised by practitioner
- barriers and facilitators to implementation raised by practitioner or observed

The observer will have a conversation with the coach before the observation about the setting context and delivery which may inform the mentoring session.

Online training and networking events

Semi-structured observation of each of the online TWiTCH champion training and online network events will be undertaken (up to four during the project) which will look at the role of these sessions within the programme, practitioner engagement with the sessions and the challenges and facilitators to delivery across settings identified in the meetings.

TWiTCH in practice

Semi-structured observation of TWiTCH practice will also be undertaken in the four case study settings during the visit towards the end of setting delivery. These observations will look at how the intervention is delivered, how closely it aligns with programme materials, and the engagement of children in the session. A guided observation schedule will be used to support the researcher in capturing each of these elements.

Observation process

The first observation of each type of session will be conducted by two researchers⁷ (the project PI and the research associate). This will allow for discussion and comparison of the observation and field notes taken, checking consistency of field notes and improving the reliability of subsequent observations.

We acknowledge that researchers visiting the settings to conduct observations is likely to change the practice of the practitioner in the setting and the coach. Observed participants are more likely to be more aware of what think they 'should' be doing than in unobserved normal practice and likely to change their practice accordingly. However, observation will still highlight the level of understanding those being observed have about programme delivery; deviations from expected procedures are likely to indicate a lack of understanding of the practitioner or coach about the process. Observation data will be triangulated with coach logbook records and with practitioner self-report data of practice to avoid over reliance on this data. Despite the potential for change in behaviour, direct observation is still valuable to see how the intervention functions in a nursery setting. Observations of online training are expected to minimise the Hawthorne effect due to the unobtrusiveness of the observer in an online meeting setting and group nature of these sessions.

Audio recordings of Story Time Practice

The coded audio recordings of Story Time Practice described in the impact evaluation will also be used in the IPE in a more exploratory way to look at the more in depth ways that practitioner behaviours vary between the intervention and control groups in the behaviours they use when sharing stories with children in small groups outside of TWiTCH. This will allow investigation of

⁷ With the exception of the first coach training session which was observed only by the project PI.

whether TWiTCH has become embedded in intervention group practice beyond the specific TWiTCH sessions and will allow comparison with usual practice in control group settings.

To do this the coded behaviours captured in the measure (see Appendix 3) will be reported descriptively to provide a comprehensive picture of practitioner story time practice across the two groups. This includes the frequency of individual prompt types, all Teaching Pedagogy items, Vocabulary Instruction items, Print Instruction items, Phonological Awareness items, total session duration, and teacher talk percentage during engagement around the story (before and after story reading). These descriptive comparisons will follow the approach used in the TWiTCH pilot report (Fitzfussell et al., 2025).

This two-tier approach using this data in both the Impact Evaluation and Implementation and Process Evaluation uses psychometrically supported ELLECCT section scores for the formal impact analysis, with the full battery of coded items reported descriptively, balances methodological rigour with the richness of data available from the coding framework. Items on the coding framework that are specific to TWiTCH pedagogy and do not map to validated ELLECCT sections (e.g., teaching pedagogy) have not undergone independent psychometric validation and are therefore treated as exploratory descriptive data rather than formal outcome measures and therefore included in the IPE.

Monitoring Implementation Fidelity

Administrative records of practitioner attendance at CPD and support sessions will be used to measure the extent to which settings complied with the training and CPD elements of the programme. Data analytics of the online training video will be used to monitor the extent to which this video is watched across the programme.

Coach logbook records which are completed by coaches after each meeting, visit and training session with a setting will be used to monitor wider implementation fidelity with the programme aiding in the capture of the information detailed in Table 6 for the impact evaluation compliance. Post-intervention survey data will also capture self-reported data as to the extent to which the programme has been completed by settings and engagement with the different elements of the programme.

The IPE will provide greater exploration of implementation fidelity using exploratory analysis of the different core and facilitative components of compliance as described in Table 6 in the IE compliance section. Champion interviews will also explore implementation fidelity in more depth and reasons for variation across a smaller number of settings.

Coded audio recordings of story time practice in intervention settings will also be used to explore the extent to which different aspects of TWiTCH practice are incorporated into shared story time practice beyond the TWiTCH project sessions.

Analysis

Survey data

Descriptive statistics will be presented for the quantitative survey questions exploring data by group allocation, setting type (PVI, Maintained) and Champion vs practitioner role in the project.

Scores from the three self-report scales of practitioner confidence, knowledge and practice measures will be analysed using similar linear models to the impact evaluation outcomes comparing post-intervention measure scores between intervention group and control group practitioners using baseline scores as covariates in the model and the inclusion of the minimisation factors. This will allow us to look at whether the intervention group report higher levels of confidence, knowledge and practice compared to control group settings testing that aspect of the logic model.

Qualitative survey responses will be coded using both deductive and inductive coding in a multi-step process as described by Bingham (2023), further described below for interview data.

Interviews

All data collected in interviews will be analysed using both deductive and inductive coding in a multi-step process as described by Bingham (2023). Transcripts will be initially coded deductively using the pre-developed codebook based on the project's research questions, theory of change and pilot evaluation previous findings to organise the data. Where data does not fit within these codes new codes will be generated. Inductive open coding will then be used within the deductive coding categories to add more nuanced coding within the pre-existing codes and to support understanding of the data. Interview data analysis will be conducted using NVivo software.

Audio recordings of Story Time Practice

Descriptive comparisons between intervention group and the control group for frequency of use of all items in the coding framework will follow the approach used in the TWiTCH pilot report (Fitzfussell et al., 2025). The composition of prompt types (the proportion that are open-ended versus closed) will be reported descriptively to examine whether TWiTCH shifted the balance toward more cognitively demanding prompt types as intended. Descriptive statistics will also be presented for the percentage of teacher time talking during the story reading and outside of reading the story book during the session for both intervention and control group practitioners.

Observation data

Observation data from training sessions and mentoring sessions will be analysed deductively to look at level of consistency in delivery of training sessions between coaches, fidelity to the programme and engagement of participants in the training/mentoring sessions.

Compliance data

Descriptive statistics of each of the items in the compliance measure will be reported to explore the variation in compliance/fidelity for the different core elements of the TWiTC program. Exploratory analysis will investigate whether these vary for different types of settings and by size of setting.

Coach logbooks for each setting will also be analysed using the same qualitative analysis - deductive and inductive coding technique described above (Bingham, 2023) looking for barriers and facilitators to implementing the programme with fidelity, adaptations made to the programme and how coaches supported settings in their delivery of the programme.

Cost evaluation design

The cost evaluation will follow the most recent guidance from the EEF (2023) using the ingredients models to identify the different components of the programme and their associated actual cost or resource cost. Costs will be calculated assuming delivery of the programme over three years, with a new cohort of children assumed each year. The cost per setting for delivery across the three years will first be established using actual costs for the first year of delivery and agreeing the expected costs for ongoing delivery for a following two years with the delivery team. To establish the cost per child for the programme, the average number of children across all participating nurseries will be used, with this number assumed to take part each year over the three years. Sensitivity analysis will be conducted using the minimum and maximum number of children in the cohort in each setting to show how cost per child varies per setting. The time required for setting staff to deliver the programme will be presented separately to the costs and described by initial set-up requirements and ongoing delivery requirements.

A Cost and Scale up workshop between the delivery team and evaluation team will explore and agree the cost for providing the TWiTC program during the delivery year and assumed costs for a further two years of delivery. The workshop will also explore future scale up cost models. The expected time that staff will spend delivering the programme during the delivery year as well as any feedback the delivery team have received about setting costs will also be covered in the workshop.

Data for the programme time, cover requirements and additional costs for settings will be collected through the usual practice surveys for practitioners and nursery senior leaders at the end of the programme and explored in interviews with TWiTC Champions and in case study settings.

Ethics and registration

Ethical approval

Ethical approval for the evaluation activities has been sought and received from the Durham University School of Education Ethics Committee on 11th March 2026. Further updates to ethical approval will be sought by the evaluation team throughout the project when new data collection

tools are drafted or if any further changes are made. Ethical approval for delivery of the TWITCH programme has been sought from the ethics board at Sheffield Hallam University by the delivery team and is currently awaiting approval.

Trial registration

The trial will be registered on the ISRCTN registration once review of the protocol is complete and the registration details will be added to the protocol.

Agreement to take part in the trial

Informed, opt-in consent will be obtained from settings, parents for child participation, and directly from adult participants for each research element.

Settings will be provided with the setting information sheet when the project is advertised initially to settings. If they are interested in taking part, they will then complete an expression of interest form and will be invited to attend a briefing event online, to watch a recording of a briefing event or to have a short one-to-one call to find out more about the project. They will be provided with a Memorandum of Understanding (MoU) which has full details of what the project requires from participants and the roles and responsibilities of those involved. The setting manager is asked to complete this to sign up to the project.

Settings will be asked to recruit children through contact with parents/carers. They will provide parents/carers with the information sheet and ask them to complete the consent form to sign their child up to the project. A short video accessed via link or QR code will provide a spoken summary of the project for parents/carers.

Settings will also pass on the email address of practitioners they want to take part in the study (a project lead/TWITCH Champion, another practitioner and a senior leader). An email will be sent to all those taking part to brief them about the project and provide a link to the baseline survey. A full information sheet will be provided at the start of the baseline survey and participants will consent to take part by ticking a box before the survey begins.

For interviews and case studies, information sheets will be sent by email to potential participants about the specific element of the project. Practitioners will be asked to indicate by email whether they consent to take part. Verbal briefing and consent will take part before the interview begins.

Data protection

All information collected as part of this project relating to participants will be treated in the strictest confidence and processed in accordance with General Data Protection Regulations (GDPR). Only data required to conduct the evaluation will be collected and all data will be collected, stored and shared securely with access only to those working on the project.

The lawful basis the evaluation team are relying on for collecting and processing data as part of this project is 'public task', which according to General Data Protection Regulations (GDPR) Article 6 (1) (e) means 'the processing is necessary for an activity being carried out as part of the

University's public task, which is defined as teaching, learning and research'. This project is carrying out research and does not intend to process special category personal data.

DU and SHU will act as data controllers or joint data controllers for the project. Navigator Insight will act as data processors to conduct child assessments. University of Manchester will act as data processor but will only process anonymised data. OxEd will act as data processor for the LanguageScreen assessment. Data sharing agreements will be put in place between SHU and DU to manage the sharing of data for the project. Nursery settings will also be asked to sign data sharing agreements with DU as there will be sharing data between DU and the setting regarding the children and practitioners in the setting. Collaboration agreements will be put in place between DU and University of Manchester, and between Navigator Insight which will set out expectations for data management and sharing. These agreements will ensure consistent ethical and data management standards.

A Data Protection Impact Assessment (DPIA) was conducted for this project by the principal investigators, describing why the collection of personal data is necessary, the data that will be collected and the ways that risks will be minimised and this has been reviewed and approved by Durham University Information Governance. The data retention period for storing and processing personal data is five years beyond the end of the project (May 2034) to allow for the verification of data for further publishing of papers beyond the report. At this point, personal data will be deleted or anonymised by the Evaluation Team. Anonymous data may be kept indefinitely and potentially shared with other research teams for future research. At the end of the evaluation study, data will be submitted to the EEF's data archive. At this point, the EEF and its archive manager (FFT Education) will become responsible for controlling and processing the evaluation study data.

A data management plan has been written by the PI and will be implemented by the evaluation team to manage the handling, storing and sharing of project data.

Participants (setting staff and parents/carers) will be provided with information sheets and privacy notices before agreeing to take part in the project which set out what, and how, data will be collected and processed as part of the project, how it would be stored, who it would be shared with along with information about participants ability to withdraw from data processing as part of the project and the retention period for project data.

Personnel

Delivery Team Personnel

Sheffield Hallam University

Fufy Demissie:	Fufy is the TWiATCH project director who developed the programme and materials and will oversee its delivery. She is responsible for training and managing TWiATCH coaches, providing champion training, and facilitating the network meetings during the project. If needed, Fufy will also serve as a TWiATCH coach during the project.
Sally Pearce	Sally is the project's strategic lead and advises on the programme development, and delivery as well as supporting the recruitment of settings.
Louisa Pinder	Louisa is the project manager responsible for managing the day-to-day delivery of the TWiATCH programme. Louisa is the main contact for settings during recruitment and delivery.
Sophie Keegan	Sophie is the project officer who will be supporting project delivery.
TWiATCH coaches	TWiATCH coaches with experience of delivering Early Years Professional Development for supporting children's language skills will provide training, mentoring, and support to each nursery setting during the project. These are likely to be employed on a consultancy basis by Sheffield Hallam University.

Evaluation Team Personnel

Durham University

Vic Menzies:	Vic is the principal investigator on the project and has overall responsibility for the project design and protocol, project delivery, and reporting and project management.
Xiaofei Qi:	Xiaofei is a co-investigator on the project and will provide advice to the team as a specialist in early years practice. She will lead the IPE, oversee IPE data collection quality assurance, and contribute to the research design.
Johny Daniel:	Johny is a co-investigator on the project and will lead on practitioner practice outcomes, including audio recordings, surveys, and observations.
Akansha Singh:	Akansha is co-investigator and lead statistician on the project who will lead on the impact evaluation statistical design and analysis.

Research Associate: This role (TBA) will involve managing the day-to-day delivery of the evaluation, drafting data collection tools, conducting data collection (observation, interviews, surveys), analysing data, and contributing to the report writing.

Junior statistician: This role (TBA) will provide support to the statistician on the writing of the SAP, cleaning the data, conducting the IE analysis and contributing to the reporting.

University of Manchester

Kelly Burgoyne: Kelly will provide specialist advice in early years language development, programme development and assessment throughout the project. She will provide Navigator Insight with training for the delivery and scoring of child assessments and scoring, will lead on assuring the quality of assessment scoring.

Navigator Insight Limited (formerly Qa Research)

Navigator Insight will manage child-level assessment delivery and data collection across nursery settings. A team of 30 assessors will be employed by Navigator Insight to collect assessment data at each assessment time point.

Katie Morris: As Head of Evaluation Assessments, Katie will be the project lead at Navigator Insight with overall responsibility for the delivery of the child assessments.

Rosie Walker-Lyne: As Senior Executive, Rosie will lead the recruitment and training of assessors and liaison with schools to schedule testing visits.

Rachel Brown: As Education Assessment Manager: Rachel will oversee the moderation and quality checking process.

Karen Tarvin: As research Manager, Karen will manage the internal assessment scheduling process.

John Lillie: As Data Analysis Manager, John will provide IT support and lead on the production of assessment data files.

Risks

Table 8. Risk management approach

Risk	Impact	Likelihood (1 - rare to 5 - very high)	Mitigation
Low nursery sign-up	Delays to starting, underpowered study, potential to abandon study	3	• ET & DT to co-host recruitment events and create accessible materials (e.g. videos) • Use of incentive payments • Include both PVI and school settings • Batch randomisation to reduce delays • Possibility to extend setting recruitment into September • Possibility to expand target regions
Insufficient children recruited	Underpowered study, delays, setting withdrawal	3	• Collect expected number of children at expression of interest to prioritise larger settings • Provide accessible parent info (incl. translations) and video • Provide resources to support settings with talking to parents • Lower minimum numbers per setting to maintain study viability • Recruitment of additional 15% of settings to form project waitlist so that setting who don't recruit minimum numbers of children can be quickly replaced.
Child data collection difficulties	Incomplete/unreliable data, inconsistent assessments	2	• Close work with Navigator Insight with regular checks to identify any issues at early stage • Quality assurance processes in place at Navigator Insight to ensure

			assessors are monitored and any issues are caught early • Hire assessors experienced with young children • Provide thorough training and include quality assurance (e.g. audio recording) • Limit sessions to 20 mins with breaks • Use shortened assessments if needed
Poor response to practitioner completed measures	Weak secondary outcome data	3	• Ensure this commitment to taking part is clear in project documentation at sign up • Use only essential, aligned measures to reduce burden • Consider use of incentives for practitioner completion of measures if low response
High attrition (settings) especially of control settings	Data loss, reduced validity	2	• Build strong relationships via regular contact • Develop communication plan early • Emphasise control group value • Support disengaged settings via calls • Encourage data submission even if intervention is stopped • Minimise data collection burden during setup • Consider offer of end-of-programme incentives for control settings
Child relocation	Data loss, reduced validity	2	• Obtain consent to follow children • Monitor missing pupils • Arrange assessments with parents if needed

Delays or difficulty with recruitment of Project Manager/Research Associate position with required skills	Delays to delivery of project and risk to overall project delivery	2	• Temporary reassignment of existing department staff member or other project team member to cover delay to recruitment for recruitment and early project work especially for May to September period when teaching is minimal • Temporary project manager could be put in place from university pool to cover for delays or absence • Potential to split Project Manager and Research Associate into two separate roles and recruit to the different skill sets should no appropriate candidate be identified for the joint role • Use of existing networks with evaluators and researchers with language development knowledge to promote role to assure it gets to appropriate candidates • 80% FTE equivalent and a contract of more than a year increases the desirability of the role attracting more suitable candidates
---	--	---	---

Timeline

Table 9. Timeline

Dates	Activity	Staff responsible/leading ⁸
Nov 25-Jan 26	Trial Set up	DU/SHU/EEF/NI
Feb 26	Ethics approval	DU
Mar 26	Submission of trial protocol	DU
Mar-Jul 26	Recruitment of nursery settings	SHU
May-Aug 26	Recruitment of assessors	NI
May, Jul, Sept 26	Observation of 'Train the trainer' coach training (4 sessions)	DU
Jul-Oct 26	Recruitment of children	DU/SHU
Aug-Sept 26	Assessor training	NI/DU
Sept-Nov 26	Baseline child assessment	NI
	Baseline practitioner and SL surveys	DU
Sept/Oct/Nov 26	Weekly rolling randomisation	DU
Nov 26-Jul 27	Programme delivery	SHU
Nov-Dec 26	Observation of practitioner training sessions	DU
Nov- Dec 26	Post training practitioner survey	DU
Dec 26- Jan 27	1 st Remote Interviews with Practitioners	DU
	1 st Coach focus groups	DU
	1 st Developer interview	DU
Jan-Jun 27	Observation of mentoring visits	DU
	Observation of Champion network meetings	DU
Apr/May 27	Case study visits	DU
	2nd Remote Interviews with Practitioners	DU

⁸ DU is Durham University, SHU is Sheffield Hallam University, EEF is Education Endowment Foundation, NI is Navigator Insight

	2nd Coach focus groups	DU
Jun - Jul 27	Endline child assessment	NI
	Storytime audio recordings	DU
	Endline practitioner and SL surveys	DU
Aug/Sept 27	2 nd developer interview	DU
	Cost and scale up workshop	DU/SHU
Sept 27 -Jan 28	Data analysis, report writing	DU
Jan 28	Submit draft report	DU
Feb-Jun 28	Review and report edits	DU/EEF/SHU
Jun 28	Submit final report	DU
Jun 28	Submit project data to EEF archive	DU
Nov 28	Apply to NPD for EYFS Profile data	DU
Jan 29-Mar 29	Analysis of EYFS Profile data	DU
Apr 29	Submit draft report addendum	DU
Apr-May 29	Review and addendum edits	DU/EEF/SHU
Jun 29	Submit final report addendum	DU

References

- Alexander, R. J. (2017) *Towards Dialogic Teaching: Rethinking Classroom Talk (5th edn)*, York: Dialogos.
- Altman, D. G., & Bland, J. M. (2005). Treatment allocation by minimisation. *BMJ*, 330(7495), 843. <https://doi.org/10.1136/bmj.330.7495.843>
- Barone, C., Chambuleyron, E., Vonnak, R., & Assirelli, G. (2019). Home-based shared book reading interventions and children's language skills: A meta-analysis of randomised controlled trials. *Educational Research and Evaluation*, 25(5-6), 270–298. <https://doi.org/10.1080/13803611.2020.1814820>
- Beitchman, J. H., Brownlie, E. B., & Bao, L. (2014). The impact of language development on psychosocial and emotional development. In P. Brooks, V. Kempe, & J. Goetz (Eds.), *Encyclopedia of Language Development* (pp. 284–286). SAGE Publications.
- Benítez-Burraco, Antonio, StefanHartmann, and MichaelPleyer. 2025. The Role of Play in Language Structure, Acquisition and Evolution. *Language and Linguistics Compass*: e70021. <https://doi.org/10.1111/lnc3.70021>.
- Bhatti, S., Lilley, C., McKaskill, M., Morgan, H., Phillips, M., Phillips, N., Poquillion, G. & Ruddick-Trentmann, J. (2024). *Hanen learning language and loving it (LLLI)*. Evaluation Report. Education Endowment Foundation.
- Bingham, A. J. (2023). From Data Management to Actionable Findings: A Five-Phase Process of Qualitative Data Analysis. *International Journal of Qualitative Methods*, 22. <https://doi.org/10.1177/16094069231183620>
- Brian, A., Merewood, J., Rosa Speciani, E., Oades, F., MacLeod, E., & Groom, M. (2025). *Using the Orchestrating Numeracy and the Executive (THE ONE) Programme to improve maths attainment, a two-armed cluster randomised trials*. Evaluation Report. Education Endowment Foundation.
- Cham, H., Reshetnyak, E., Rosenfeld, B., & Breitbart, W. (2017). Full information maximum likelihood estimation for latent variable interactions with incomplete indicators . *Multivariate Behav Res.*, 52, 12–30. doi:10.1080/00273171.2016.1245600
- Curenton, S., Sims, J., Rochester, S., & Gardner, B. (2018). The Conversation Compass© Communication Screener-Revised. *Early Childhood Research Quarterly*, 47, 182-193.
- Demissie, F. (2019) 'TWITCH (Thinking With Tales for Children) Project Report, (unpublished).
- Department for Education (2025). *Early years foundation stage profile handbook*. <https://www.gov.uk/government/publications/early-years-foundation-stage-profile-handbook/early-years-foundation-stage-profile-handbook> accessed 27th March 2026.
- Department for Education (2026). *Early years foundation stage assessment support*. <https://www.gov.uk/guidance/early-years-foundation-stage-profile-assessment-support> accessed 28th April 2026

Dowdall, N., Melendez-Torres, G.J., Murray, L., Gardener, F., Hartford, L., & Cooper, P. (2020). Shared picture book reading interventions for child language development: A systematic review and meta-analysis, *Child Development*, 91, 383-399. doi: 10.1111/cdev.13225

Doyle, B. G., & Bramwell, W. (2006). Promoting emergent literacy and social–emotional learning through dialogic reading. *The Reading Teacher*, 59(6), 554-564.

Eadie, P., Bavin, E., Bretherton, L., Cook, F., Gold, L. and Reilly, S. (2021) ‘Predictors in Infancy for Language and Academic Outcomes at 11 Years’, *Pediatrics*, 147 (2), e20201712. DOI: 10.1542/peds.2020-1712

Education Endowment Foundation (n.d.) *Early years measures database*.

<https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluation/eef-outcome-measures-and-databases/early-years-measures-database-2/early-years-measures-database> accessed 27th March 2026.

Fitzfussell, N., Menzies, V., Eerola, P., Daniel, J., Burgoyne, K., Qi, X., & Cramman, H. (2025). *Talk With Tales for Children (TWITCH) programme. Pilot evaluation report*. Education Endowment Foundation.

Flack, Z. M., Field, A. P., & Horst, J. S. (2018). The effects of shared storybook reading on word learning: A meta-analysis. *Developmental psychology*, 54(7), 1334

Gago-Galvagno, L. G., Elgier, A. M., Tabullo, A. J., Huaire-Inacio, E. J., Herrera-Alvarez, A. M., Zambrano-Villalba, C., ... & Miller, S. E. (2025). Use of screens, books and adults’ interactions on toddler’s language and motor skills: A cross-cultural study among 19 Latin American countries from different SES. *Plos one*, 20(2), e0314569.

Gorard, S., Siddiqui, N. and See, B. H. (2015) *Philosophy for Children: Evaluation Report and Executive Summary*, London: Education Endowment Foundation.

Hargrave, A. C., & Sénéchal, M. (2000). A book reading intervention with preschool children who have limited vocabularies: The benefits of regular reading and dialogic reading. *Early Childhood Research Quarterly*, 15(1), 75-90.

Haynes, J. & Murris, K. (2012). *Picturebooks, pedagogy and philosophy*. Routledge

Horst, J. S., Parsons, K. L., & Bryan, N. M. (2011). Get the story straight: Contextual repetition promotes word learning from storybooks. *Frontiers in Psychology*, 2, 17.

Hedges, L. V. (2007). Effect sizes in education research. *Educational Researcher*, 36(3), 148–160. <https://doi.org/10.3102/0013189X07307795>

Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis*, 29(1), 60–87. <https://doi.org/10.3102/0162373707299706>

- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Hughes, C., Daly, I., Foley, S., White, N. and Devine, R. T. (2015). Measuring the foundations of school readiness: Introducing a new questionnaire for teachers – The Brief Early Skills and Support Index (BESSI). *British Journal of Educational Psychology*, 85, 332- 356.
- Hulme, C., Stothard, S., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., & Snowling, M. (2011). *York Assessment of Reading for Comprehension*. Early Reading. GL Assessment.
- Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. V. (2010). Sources of variability in children’s language growth. *Cognitive psychology*, 61(4), 343-365. –
- Jay, T., Willis, B., Thomas, P., Taylor, R., Moore, N., Burnett, C., Merchant, G. and Stevens, A. (2017) *Dialogic Teaching: Evaluation Report and Executive Summary*, London: Education Endowment Foundation.
https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/Dialogic_Teaching_Evaluation_Report.pdf?v=1734707307.
- Kalland, M., & Linnavalli, T. (2023). Associations between social-emotional and language development in preschool children. Results from a study testing the rationale for an intervention. *Scandinavian Journal of Educational Research*, 67(5), 791-804.
- Kline, R. B. (1998). *Principles and practice of structural equation modeling*. Guilford Press.
- Law, J., Charlton, J. & Asmussen, K. (2017). *Language as a child wellbeing indicator*. Early intervention foundation.
- Law, J., Rush, R., Parsons, S., and Schoon, I. (2009). Modelling developmental language difficulties from school entry into adulthood: Literacy, mental health and employment outcomes. *Journal of Speech, Language and Hearing Research* 52, pp. 1401-16
- Lindorff, A., Sylva, K., Ereky-Stevens, K. and Joseph, A. (2022) *Coaching Early Conversation, Interaction and Language (CECIL): Impact Evaluation*, London: The Sutton Trust.
<https://www.suttontrust.com/wp-content/uploads/2022/03/CECIL-Evaluation-Oxford-14.02.2022.pdf>
- Lipscomb, S. T., Chandler, K. D., Abshire, C., Jaramillo, J., & Kothari, B. (2022a). Early childhood teachers’ self-efficacy and professional support predict work engagement. *Early Childhood Education Journal*, 50(4), 675–685. doi:10.1007/s10643-021-01182-5
- Lord, P., Dirie, A., Kettlewell, K. and Styles, B. (2021) ‘*Evaluation of Philosophy for Children—An Effectiveness Trial: Evaluation Report*’, London: Education Endowment Foundation.
- Massaroni, V., Delle Donne, V., Marra, C., Arcangeli, V., & Chieffo, D. P. R. (2024). The relationship between language and technology: How screen time affects language development in early life-A systematic review. *Brain Sciences*, 14(1), 27.
doi:10.3390/brainsci14010027

- McGee, L. M., & Schickedanz, J. A. (2007). Repeated interactive read-alouds in preschool and kindergarten. *The reading teacher*, 60(8), 742-751.
- Menzies, V., Cramman, H., Eerola, P., Hugill-Jones, J., Akhter, N. & Einbeck, J. (2022). *Parents and Children Together (PACT). Evaluation Report*. Education Endowment Foundation. <https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/PACT-Final-Report.pdf?v=1774557580>
- Menzies, V., Eerola, P., Ashraf, B., Zhang, Q., Cramman, H., & Einbeck, J. (2024). *Evaluating the impact of the Parents and Children Together (PACT) programme on the language skills of 3-4-year-old nursery children. A two-armed randomised trial. Evaluation Report*. Education Endowment Foundation. (<https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/pact-parents-and-children-together-regrant>)
- Merrell, C., Little, J. and Coe, R. (2014) 'Is the Attainment Gap among Primary Aged Children Decreasing?', in C. Wood and R. Scott (eds), *Harnessing What Works in Eliminating Educational Disadvantage: A Tale of Two Classrooms*, London: Demos (pp. 45–52). https://dera.ioe.ac.uk/23193/1/Two_classrooms_-_web.pdf
- Mihai, A., & Butera, G. (2019). " Let's Read this Book Again": Using Repeated Reading to Support Literal and Inferential Language Skills in Inclusive Preschool Classrooms. *The Dialog: A Journal for Inclusive Early Childhood Professionals*, 22(2).
- Miller, S., Dunne, L., Smith, A. & Laishley, A. (2023). *Peep learning together programme. Evaluation report*. Education Endowment Foundation.
- Mol, S. E., Bus, A. G. and de Jong, M. T. (2009) 'Interactive Book Reading in Early Education: A Tool to Stimulate Print Knowledge As Well As Oral Language', *Review of Educational Research*, 79 (2), pp. 979–1007.
- Noble, C., Sala, G., Peter, M., Lingwood, J., Rowland, C., Gobet, F. & Pine, J. (2019). The impact of shared booked reading on children's language skills: A meta-analysis. *Education Research Review*, 28, <https://doi.org/10.1016/j.edurev.2019.100290>
- Oxford University Press. (2018). *Why closing the word gap matters*: Oxford language report.
- Piasta, S., Justice, L., Cabell, S., Wiggins, A. and Turnbull, K. (2012) 'Impact of Professional Development on Preschool Teachers' Conversational Responsivity and Children's Linguistic Productivity and Complexity', *Early Childhood Research Quarterly*, 27, pp. 387–400. <https://doi.org/10.1016/j.ecresq.2012.01.001>
- Pillinger, C., & Vardy, E. J. (2022). The story so far: A systematic review of the dialogic reading literature. *Journal of Research in Reading*, 45(4), 533-548.
- Pankratz, Mary & Plante, Elena & Vance, Rebecca & Insalaco, Deborah. (2007). The Diagnostic and Predictive Validity of The Renfrew Bus Story. *Language, Speech, and Hearing Services in Schools*. 38. 390-399. 10.1044/0161-1461(2007/040).

Purdon, A. (2016). Sustained shared thinking in an early childhood setting: An exploration of practitioners' perspectives. *Education 3-13*, 44(3), 269–282.
doi:10.1080/03004279.2014.907819

Renfrew, C. (2010). *Bus Story Test: Revised Edition (The Renfrew Language Scales)*, Speechmark.

Ryan, R., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness (1st ed.)*. New York: Guilford Press.

Robinson-Smith, L., Fairhurst, C., Stone, G., Bell, K., Elliott, L., Gascoine, L., Hallett, S., Hewitt, C., Hugill, J., Torgerson, C., Torgerson, D., Menzies, V., Ainsworth, H. (2018). *Maths Champions. Evaluation report and executive summary*. Education Endowment Foundation.

Robinson-Smith, L., Menzies, V., Cramman, H., Wang, Y., Fairhurst, C., Hallett, S., Beckmann, N., Merrell, C., Torgerson, C., Stothard, S., & Siddiqui, N. (2019). *Easypeasy: Learning through play. Evaluation report*. Education Endowment Foundation.

Robinson-Smith, L., Whiteside, K., Torgerson, C., Fairhurst, C., Podmore, D., Davill, T., Bell, K., Qi, X., Leggett, H., Elliott, L., Hewitt, C., Baird, K., Menzies, V., Torgerson, D., Ainsworth, H. (2024). *Independent evaluation of Maths Champions in nursery to develop children's early numeracy: A two-armed cluster randomised controlled trial. Evaluation Report*. Education Endowment Foundation.

<https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/Evaluation-report-amended.pdf?v=1759339603>

Sammons, P., Elliot, K., Sylva, K., Melhuish, E., Siraj-Blatchford, I. and Taggart, B. (2004) 'The Impact of Pre-School on Young Children's Cognitive Attainments at Entry to Reception', *British Educational Research Journal*, 30, pp. 691–712.
<https://doi.org/10.1080/0141192042000234656>

Sammons, P., Gu, Q., Day, C. and Ko, J. (2011) 'Exploring the Impact of School Leadership on Pupil Outcomes: Results from a Study of Academically Improved and Effective Schools in England', *International Journal of Educational Management*, 25, pp. 83–101.

Saghaei, M., & Saghaei, S. (2011). MinimPy: A software for minimisation in clinical trials. *Journal of Medical Signals and Sensors*, 1(1), 55–58.

Siraj-Blatchford, I. and Manni, L. (2009) "“Would you like to tidy up now?”: An Analysis of Adult Questioning in the English Foundation Stage', *Early Years*, 28, pp. 5–22.

Siraj-Blatchford, I., Muttock, S., Sylva, K., Gilden, R., & Bell, D. (2002). *Researching effective pedagogy in the early years* (Vol. 356). London: Department for Education and Skills.

Snow, C. E., Burns, M. S. and Griffin, P. (1998) *Preventing Reading Difficulties in Young Children*, Washington, DC: National Academy Press.

Stanley, S. (2012). *Why think?: philosophical play from 3-11*. Bloomsbury Publishing.

Subosa, M., Oades, F., Dysart, E., Tracey, L., Brian, A., & Speciani, E.R. (2025), *Concept Cat: A two-armed cluster randomised controlled trial. Evaluation report*. Education Endowment Foundation.

Suffill, E., Schonberg, C., Vlach, H. A., & Lupyan, G. (2022). Children's knowledge of superordinate words predicts subsequent inductive reasoning. *Journal of Experimental Child Psychology*, 221, 105449.

Sun, Y. J., Sahakian, B. J., Langley, C., Yang, A., Jiang, Y., Kang, J., ... & Feng, J. (2024). Early-initiated childhood reading for pleasure: associations with better cognitive performance, mental well-being and brain structure in young adolescence. *Psychological Medicine*, 54(2), 359-373.

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I. and Taggart, B. (2004) *The Effective Provision of Pre-School Education (EPPE) Project Technical Paper 12: The Final Report-Effective Pre-School Education*, London: IOE.

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2004). *The effective provision of pre-school education (EPPE) project: Findings from pre-school to end of key Stage 1 research brief*.

Tsomokos, D. I., & Raviv, L. (2024). A bidirectional association between language development and prosocial behaviour in childhood: Evidence from a longitudinal birth cohort in the United Kingdom. *Developmental Psychology*. <https://doi.org/10.1037/dev0001730>

Tymms, P., Merrell, C., Hawker, D. and Nicholson, F. (2014) *Performance Indicators in Primary Schools: A Comparison of Performance on Entry to School and the Progress Made in the First Year in England and Four Other Jurisdictions*, London: Department for Education. <https://www.gov.uk/government/publications/performance-indicators-in-primary-schools>

van der Wilt, F., Bouwer, R. and van der Veen, C. (2022) 'Dialogic Classroom Talk in Early Childhood Education: The Effect on Language Skills and Social Competence', *Learning and Instruction*, 77, 101522. <https://doi.org/10.1016/j.learninstruc.2021.101522>

van der Wilt, F., Smits-van der Nat, M., & van der Veen, C. (2022). Shared Book Reading in Early Childhood Education: Effect of Two Approaches on Children's Language Competence, Story Comprehension, and Causal Reasoning. *Journal of Research in Childhood Education*, 36(4), 592–610. <https://doi.org/10.1080/02568543.2022.2026540>

Wartenberg, T. E. (2022). *Thinking Through Stories: Children, Philosophy, and Picture Books*. Routledge.

Weadman, T., Serry, T. and Snow, P. C. (2022) 'The Development and Psychometric Properties of a Shared Book Reading Observational Tool: The Emergent Literacy and Language Early Childhood Checklist for Teachers (ELLECCT)', *First Language*, advance online publication: <https://doi.org/10.1177/01427237211056735>

West, G., Lervåg, A., Birchenough, J.M.H., Korell, C., Rios Diaz, M., Duta, M., Cripps, D., Gardner, R., Fairhurst, C. and Hulme, C. (2024), Oral language enrichment in preschool

improves children's language skills: a cluster randomised controlled trial. *Journal of Child Psychology and Psychiatry*, 65: 1087-1097. <https://doi.org/10.1111/jcpp.13947>

West, G., Birchenough, J., Korell, C., Rios Diaz, M., Lervåg, A., Duta, M. Cripps, D., Gardner, R., Fairhurst, C. & Hulme, C. (2023). *An efficacy trial of the Nuffield Early Language Intervention in Preschool (NELI Preschool)*. Final report to the Nuffield Foundation.

https://www.education.ox.ac.uk/wp-content/uploads/2018/06/NELI-PRESCHOOL_FINAL-REPORT_23_11_15.pdf

West, G., Snowling, M., Lervåg, A., Buchanan-Worster, E., Duta, M., Hall, A., McLachlan, H. and Hulme, C. (2021) Early Language Screening and Intervention Can Be Delivered Successfully at Scale: Evidence from a Cluster Randomized Controlled Trial, *Child Psychology and Psychiatry*: <https://doi.org/10.1111/jcpp.13415>

Whitehurst, G. J., Falco, F. L., Lonigan, C. J., Fischel, J. E., DeBaryshe, B. D., ValdezMenchaca, M. C., & Caulfield, M. (1988). Accelerating language development through picture book reading. *Developmental psychology*, 24(4), 552.

Appendix A: Changes since the previous EEF evaluation

Appendix Table 1: Changes since the previous evaluation

Feature		Pilot to efficacy stage
Intervention	Intervention content	Addition of short online training video introducing TWiTeach and theory before first training session. Changes to week 2 content with simplified activity format, further resources to support delivery, use of examples rather than script. Greater emphasis in training on identifying and supporting children who need additional input during week 3.
	Delivery model	All coaches now trained together through 3 coach training sessions and regular coach and developer meetings aim to increase consistency of training and support delivery. Second in-person training now takes place after settings have implemented one TWiTeach cycle. Maximum group size for delivery to children is now 10 (advised as 5-8 for pilot). Hybrid support model for coaching visits- first in person but follow up meetings can be online if agreed between coach and practitioner. All were in person for pilot.
	Intervention duration	Pilot specified 10 cycles of TWiTeach (30 weeks) with practitioners planning a cycle with their own choice of book for the 10th cycle. Due to inclusion of a pre-test, intervention shorted to 9 cycles of TWiTeach excluding 10th cycle for efficacy stage.
Evaluation	Eligibility criteria	Specified minimum of 2 early years practitioners to take part from each setting. Explicitly specified that all children should be involved in TWiTeach programme. Wider recruitment area.
	Level of randomisation	Not applicable to pilots.
	Outcomes and baseline	Not applicable to pilots.
	Control condition	Not applicable to pilots.

Appendix B: Theory of Change Assumptions Logs

Appendix table 2. Theory of change contextual assumptions

Assumptions Log 1: Contextual Assumptions

#	Contextual Assumption	Assumption Strength ⁹	Assumption Risk ¹⁰
1	Sufficient interest from settings to meet recruitment targets	Green	Red
2	Practitioners are able and motivated to take part in full programme	Amber/Green	Red
3	Following training sessions, practitioners are able to apply TWiTCH in setting	Amber/Green	Amber/Red
4	Intervention is delivered to children aged 3-4	Amber/Green	Amber/Red
6	Practitioners lack confidence in supporting young children's language and sustained shared thinking	Green	Amber/Green
7	Ratios within settings will support the group work aspect of TWiTCH	Amber/Green	Amber/Red
8	Settings are able to accommodate the coaching visits	Amber/Green	Amber/Red
9	Settings are able to release staff to attend both training sessions	Amber/Red	Red
10	Training of coaches means that all coaches are confident and skilled to deliver TWiTCH training and support settings	Green	Amber/Red
11	Intervention is specified well enough that coaches deliver training and support in a standard way	Green	Amber/Green
12	TWiTCH champion will be able to fit role into their usual workload or will be released to do this role	Amber/Green	Amber/Green

⁹ Assumption strength - How strong do you think your assumption is where the programme is delivered? Green – This assumption will hold in the vast majority of circumstances, Green/Amber – This assumption will hold in most of the circumstances, Red/Amber – This assumption will often not hold in the circumstances, Red – There is a good chance of this assumption not holding / do not know whether this assumption will hold or not

¹⁰ Assumption risk - If this assumption does not hold, how much of an impact on programme delivery with fidelity to the original design? Green – The programme could continue to be delivered with very minor impact, Green/Amber – The programme could continue to be delivered, but the impact would be substantial, Red/Amber – The programme could continue to be delivered, but without fidelity to original design, Red – The programme could not be delivered

14	staffing will remain stable across the academic year	Amber/Green	Amber/Green
15	settings have the space and routines that enables them to reach all the children	Amber/Green	Amber/Green

Appendix table 3. Theory of change causal assumptions

Assumptions Log 2: Causal Assumptions

#	Assumption / Causal Mechanism	Where in the ToC does the assumption apply?		Evidence	Evidence Strength
1	Practitioners' knowledge of dialogic book sharing is improved due to a structured programme, coach support that builds knowledge	Input/Activity	Output	Through TWiTCH training, practitioners will gain confidence, knowledge and skills in shaping interactions with young children through high quality professional development (Piasta 2012, EEF 2021; Fitzfussell et al., 2025). The programme draws on the EEF's recommendation on effective professional development: the structured programme builds professional knowledge, the experiential nature of the training and agency teachers have in selecting resources motivates them, and the programme is designed to fit in with the existing context of EY classrooms where storybook are a key feature.	Amber/Green
2	Practitioners' skills in supporting young children's language and reasoning are improved through opportunities to practice in context of usual early years practice, supported by coaching	Output	Short-term outcome		Amber/Green
3	Practitioners sense of the importance of repetition of dialogic book sharing is improved by training sessions that identify why dialogic book sharing is important and reinforced through coaching and experiential nature of training	Short-term outcome	Short-term outcome		Amber/Green
4	Practitioners are motivated to change their practice by training sessions that identify why dialogic book sharing is important and reinforced through coaching and experiential nature of training	Short-term outcome	Long-term outcome		Amber/Green

5	Practitioners are able to change their practice to incorporate dialogic story-telling and dialogic interactions	Output	Long-term outcome		Amber/Green
6	Dialogic opportunities and practitioners modelling language structures including subordinate terms, leads to improvement in children's language skills and vocabulary including children using more complex language structures to express their thoughts including 'why, because' and superordinate terms 'animals' versus dog/cat,	Output	Long-term outcome	The quality and quantity of adult-child interactions in the early years has a significant impact on children's language development (Siraj-Blatchford et al. 2002), and TWiTCH provides daily opportunities for high quality interactions. The focus on books and language development reflects EEF's key recommendations about the importance of high-quality interactions through a range of activities (Sylva et al., 2004). Dialogic reading and sharing of storybooks impacts positively on children's word-learning (Flack et al 2018) and that repeated exposure to the same stories promotes children's language skills (Mihai & Butera, 2019). There is promising early evidence from a small pilot study delivered in one nursery school initially and then rolled out to 6 schools. The original nursery school, which is in an area of disadvantage, successfully implemented TWiTCH and two of the other schools continue to deliver it and have reported a positive impact on the complexity of children's language. (Fitzfussell et al., 2025), Huttenlocher et al., (2010)- suggest more complex language by caregivers leads to more complex syntax by children and Suffill, et al.(2022) suggest that children's knowledge of superordinate words predicts subsequent inductive reasoning.	Amber/Green
7	Richer language environment in settings leads to 'children's increased confidence and skills in communication and language	Long-term outcome	Long-term outcome	Reading stories out loud to children with limited vocabulary in an early years setting on its own does not sufficiently improve language development. shared book reading impact on conceptual knowledge, inferential language skills and reading for pleasure, seen as predictive of future success (Sun et al., 2024; Whitehurst et al., (1988) irrespective of socio-economic backgrounds (Sun et al., 2024). However dialogic book sharing, as in TWiTCH, does have an impact (Hargrave and Senechal 2000).	Amber/Green
8	Improved children's language development is reflected in increased EYFS learning goal scores in communication language and literacy	Long-term outcome	Long-term outcome		Amber/Green
9	Giving practitioners agency and a certain level of choice in how to implement the programme makes it more likely that they will integrate the programme into their usual practice	Output	Short-term outcome	Ryan, R., & Deci, E. L. (2017). Self-determination theory: Basic psychological needs in motivation, development, and wellness (1st ed.). New York: Guilford Press . Lipscomb, S. T., Chandler, K. D., Abshire, C., Jaramillo, J., & Kothari, B. (2022a). Early childhood teachers' self-efficacy and professional support	Amber/Green

				predict work engagement. <i>Early Childhood Education Journal</i> , 50(4), 675–685. doi:10.1007/s10643-021-01182-5	
10	Delivery of 3-week repeated reading cycle will motivate and engage children in TWITCH sessions and lead to improved engagement with stories and vocabulary	output	Short-term outcome	TWITCH pilot evaluation (Fitzfussell et al. 2025) Mihai, A., & Butera, G. (2019). The effects of a whole-classroom repeated book reading intervention on the language development of children in Head Start. <i>HS Dialog: The Research to Practice Journal for the Early Childhood Field</i>, 22(2). Pillinger, C., & Vardy, E. J. (2022). The story so far: A systematic review of the dialogic reading literature. <i>Journal of Research in Reading</i>, 45(4), 533-548. McGee, L. M., & Schickedanz, J. A. (2007). Repeated interactive read-alouds in preschool and kindergarten. <i>The reading teacher</i>, 60(8), 742-751.	amber/green
12	Delivery of 3-week repeated reading cycle will improve children's confidence to articulate their thoughts and opinions	output	Short-term outcome	van der Wilt, F., Smits-van der Nat, M., & van der Veen, C. (2022). Shared Book Reading in Early Childhood Education: Effect of Two Approaches on Children's Language Competence, Story Comprehension, and Causal Reasoning. <i>Journal of Research in Childhood Education</i>, 36(4), 592–610. https://doi.org/10.1080/02568543.2022.2026540 Wartenberg, T. E. (2022). <i>Thinking Through Stories: Children, Philosophy, and Picture Books</i>. Routledge. Whitehurst, G. J., Falco, F. L., Lonigan, C. J., Fischel, J. E., DeBaryshe, B. D., Valdez-Menchaca, M. C., & Caulfield, M. (1988). Accelerating language development through picture book reading. <i>Developmental psychology</i>, 24(4), 552.	amber/green

13	Coach training on dialogic reading principles and ongoing support will improve practitioners' skills in dialogic reading	output	Short-term outcome	Doyle, B. G., & Bramwell, W. (2006). Promoting emergent literacy and social-emotional learning through dialogic reading. <i>The Reading Teacher</i> , 59(6), 554-564. Whitehurst, G. J., Falco, F. L., Lonigan, C. J., Fischel, J. E., DeBaryshe, B. D., Valdez-Menchaca, M. C., & Caulfield, M. (1988). Accelerating language development through picture book reading. <i>Developmental psychology</i> , 24(4), 552	amber/green
14	coach training on dialogic reading principles and ongoing support will improve practitioners' facilitative questioning using RCRI strategies (Repeat, Check, Reason, Imagine)	output	Short-term outcome	TWITCH pilot evaluation (Fitzfussell et al. 2025) Piasta, (2012) 'Impact of professional development on preschool early years practitioners' alters' conversational responsivity and children's linguistic productivity and complexity', <i>Early Childhood Research Quarterly</i> , 27 (3). Jay, T., et al. (2017). Dialogic teaching: Evaluation report and executive summary.	amber/green
15	Traditional tales are used because their relative familiarity to children, families and practitioners makes them more accessible. Importantly, traditional tales also contain contested concepts (e.g. whether Goldilocks made good choices or bad choices when she ate baby bear's porridge) that offer opportunities for children to make choices, give reasons and reflect on their views.	Input/Activity	Short-term outcome	Haynes, J. & Murris, K. (2012). <i>Picturebooks, pedagogy and philosophy</i> . Routledge	Amber/Green
16	Coaching observation of practitioners practice along with providing feedback provides opportunity for practitioner reflection and systematic improvement of practice	Input/Activity	Short-term outcome	Piasta, (2012) 'Impact of professional development on preschool early years practitioners alters' conversational responsivity and children's linguistic productivity and complexity', <i>Early Childhood Research Quarterly</i> , 27 (3).	Amber/Green
17	Continuous provision activities in week 3 enables children to further embed their knowledge of story through their play and develop their expressive and receptive language skills	Input/Activity	Short-term outcome	Benítez-Burraco, Antonio, StefanHartmann, and MichaelPleyer. 2025. "The Role of Play in Language Structure, Acquisition and Evolution." <i>Language and Linguistics Compass</i> : e70021. https://doi.org/10.1111/lnc3.70021 .	Amber/Green

18	Use of language games in week 2 that involves children expressing their views on contestable ideas/concepts within the story develops their expressive language and critical and creative thinking.	Input/Activity	Short-term outcome	Stanley, S. (2012). Why think?: philosophical play from 3-11. Bloomsbury Publishing.	Amber/Green
----	---	----------------	--------------------	--	-------------

Appendix C: Audio Recording Coding Tool

TWITCH Observation Tool:

Recording ID: _____
 Setting Name: _____ Teacher: _____
 Session time: _____
 Coder Name: _____ Coding Date: _____
 Is this for double coding: Yes

Time Story Reading Starts: _____

Item	Frequency	Time stamp
Prompts		
- Open ended prompts (from ELLECT)		
- Completion and close ended prompts (from ELLECT)		
- Distancing prompts (from ELLECT)		
- Modelling response to a prompt (with or without a puppet, for example a rabbit)		
Teaching Pedagogy		
- Reviews key ideas		
- Describes story images to students		
- Use of other visual cue		
- Repeats previously presented information		
- Encourages children to reflect on their answers		
- Repeat child's response		
- Provides students the opportunity to retell the story		
- Provides students the opportunity to role play story characters		
- Provides corrective feedback		
- Teacher ignores child		
Print Instruction		
- Select and stress a word (from ELLECT)		
- Engages students in print concepts (from ELLECT)		
- Alphabet knowledge (from ELLECT)		
Phonological Awareness		
- Syllables (from ELLECT)		
- Rhyming (from ELLECT)		
- Alliteration (from ELLECT)		
- Sound awareness (from ELLECT)		
Vocabulary Instruction		
- Provides a definition of vocabulary word (from ELLECT)		
- Provides an example and non-example of key vocabulary word		
- Engages students in extension activity about key vocabulary word		
Teacher Talk Duration		
- Duration of teacher talking during story reading		
- Teacher talk time percentage during story reading		
- Duration of teacher talking during engagement related to the story		
- Teacher talk time percentage during engagement related to the story		