Efficacy Trial of *Talking Time*©, an oral language intervention for early years Evaluation Protocol

Evaluator: NIESR
Principal investigator: Edoardo Masset

Education Endowment Foundation

Evaluation summary

Project title	Efficacy Trial of ***Talking Time***©, an oral language intervention for early years
Developer *(Institution)*	IOE, UCL's Faculty of Education and Society and University of Oxford
Evaluator *(Institution)*	National Institute of Economic and Social Research (NIESR)
Principal investigator(s)	Edoardo Masset
Protocol author(s)	Edoardo Masset, Lucy Stokes, Janine Boshoff, Carol Vincent, Ekaterina Aleynikova, Richard Dorsett
Trial design	Two-arm cluster randomised controlled trial with random allocation at setting level
Trial type	Efficacy
Pupil age range and Key stage	Age 3-4, Key Stage – Early Years
Number of settings *(at design stage)*	130
Number of children *(at design stage)*	2600
Primary outcome measure and source	Oral language skills (a composite index of Expressive vocabulary measured by the Renfrew Expressive Vocabulary Test (Renfrew, 2023) and of Information and grammar measured by the Renfrew Action Picture Test (Renfrew, 2019)).
Secondary outcome measure and source	1. Expressive vocabulary, from the Renfrew Expressive Vocabulary test (Renfrew, 2023). 2. Information, from the Renfrew Action Picture Test) (Renfrew, 2019) 3. Grammar, from the Renfrew Action Picture Test) (Renfrew, 2019) 4. Sentence repetition from the Grammar and Phonology Screening (GAPS) (Gardner et al., 2006)

Protocol version history

Version	Date	Reason for revision
1.1	September 2025	After analysis of the baseline data we decided to change the primary outcome because the GAPS assessment (which had originally been set as the primary outcome of the study) was administered with great difficulties in nurseries and delivered a large proportion (>40%) of zero and missing scores.
1.0 [***original***]	July 2024	N/A

Contents

Tables

Figures

Study rationale and background

Children from economically disadvantaged households experience disproportionate language learning delays in comparison to their less disadvantaged peers (ICAN and RCSLT, 2018). Children from poorer socioeconomic backgrounds are more than twice as likely to have speech, language, and communication needs than their more advantaged peers (Law, Charlton and & Asmussen, 2017). The recent COVID-19 pandemic and successive lockdowns have exacerbated the speech and language development gap among children from deprived socioeconomic backgrounds (González and & Bonal, 2021). A recent report by Ofsted (2020) highlighted the ongoing gaps in children's communication and language development following the pandemic, and noted the importance of early years providers offering a language-rich environment to support the development of these skills. Interventions targeted at early years settings could provide crucial support to children who are lagging behind their peers in their oral language skills.

The development of language and communication skills can have a significant impact on children's academic achievement, and their employment outcomes as adults (Roulstone et al., 2010; Education Endowment Foundation, 2018). Children with low entry levels of oral language skills are at risk of significant difficulties throughout schooling (Shanahan, 2006). Pre-school language development has been found to explain most of the effect of socio-economic status on primary school performance (Durham et al., 2007). Such difficulties can have long-lasting effects on literacy, mental health, and employment in adulthood (Law et al., 2009).

Talking Time© is a universal intervention targeting the oral language skills of children aged between 3 to 5 years. The intervention is designed to provide high quality professional development to early years practitioners that will enable them to deliver a project of engaging structured activities to support the oral language skills of children in small groups in the nursery setting. The programme aims to develop supportive conditions for language development, informed by research evidence - specifically, meaningful and engaging structured small-group experiences (Wasik, 2008), which allow children to hear and rehearse language in the context of multi-turn conversations (Romeo et al., 2018; Rowe and & Snow, 2020).

The language skills of children from disadvantaged families can be challenged in various ways. First, some children may experience a 'word gap' (Golinkoff et al., 2019) and are more likely to enter school with a reduced vocabulary in comparison to their peers (Huttenlocher et al., 2010; Hoff, 2014). Second, they may be affected by a 'grammar gap' (Pace et al., 2017) and may have a grammatical competency below what is expected for their age. Finally, there is evidence suggesting that disadvantaged children are more likely to struggle in acquiring new linguistic concepts because of poorer language processing skills (Levine et al., 2020; Weiler and & Decker, 2022).

The evidence suggests that sensitive and responsive adult-child oral language interactions can improve children's language skills by increasing the diversity and complexity of language in early years, by talking "with" children rather than "to" children, and through a gradual transition from contextualised to decontextualised conversations (Rowe and & Snow, 2020; Rowe, 2022). Such interactions should occur within small groups in supportive learning environments (Morra Pellegrino and & Scopesi, 1990; Hassinger-Das et al., 2017), and should

be guided by trained staff to support children's language growth (Dockrell *et al.*, 2017). Yet not all English early year settings operating in disadvantaged areas conduct activities conducive to oral language development, such as interactive book reading, language supporting interactions, and encouraging children to talk with each other (Wright et al., 2020; Dobinson et al. (in press)). The problem is compounded in some settings by a lack of training and experience among educators around oral language (Dockrell *et al.*, 2017).

The *Talking Time*© programme activities and manuals were originally developed at the Institute of Education (IOE) University College London (UCL) (Dockrell, Stuart and & King, 2010). Subsequently a more comprehensive manual (updated and expanded to include core texts and flexible activity prompts and plans), and a comprehensive professional development programme, were developed by the University of Oxford. These were integrated into the programme, in collaboration with IOE *Talking Time*© was first implemented in a quasi-experimental study in three nursery settings in London, in a borough ranked the third most deprived in the country and in which the population belonged to different ethnic groups (Dockrell, Stuart and & King, 2010). The study showed significant changes in vocabulary, oral comprehension, and on sentence repetition, but not on narrative skills. These results, although based on a quasi-experimental design, suggested that the programme could lead to significant improvements in children language skills. However, the lack of impact on narrative skills pointed to the need for conducting a larger experimental study employing a wider range of outcome measures to understand what specific language skills are improved.

The expanded 'integrated' programme (comprising activities for children and professional development for staff) was evaluated via randomised controlled trial involving 40 settings assessing a wide range of language learning outcomes was funded by the Nuffield Foundation in 2019-20. However, this trial was stopped during the Covid-19 pandemic.

A retrial of *Talking Time*© was funded by the Nuffield Foundation in 2021-22, involving 36 settings (Dockrell *et al.*, 2023). The impact of the intervention was assessed in terms of i) child-level outcomes for expressive vocabulary, grammar, comprehension, and oral narrative skills; ii) improved adult-child interaction at the practitioner level, and iii) improved parent/carer-report of language and behaviour. The intervention had a significant positive impact on targeted expressive vocabulary embedded within the activities. It is important to note that although the words were targeted by the intervention, practitioners were not aware of which words were being targeted. No statistically significant impacts were identified for the other language assessments, although a positive trend for sentence repetition was observed. This trial was also impacted by the COVID-19 pandemic, through staff and children's absences.

The evaluation of the Nuffield-funded trial concluded that further research would be needed a) to assess the effectiveness of the intervention in ideal conditions not affected by disruptive events such as COVID-19, b) to estimate impacts using a larger sample size, c) to allow a longer intervention period between baseline and follow up to allow the intervention to be fully embedded in settings, and d) to include an analysis of the association between the exposure to the project (quantity and quality of child-adult interactions) and the effectiveness of the intervention (Dockrell *et al.*, 2023).

The EEF is funding an efficacy trial of the *Talking Time*© programme to provide evidence for what works to support oral language development of young children in disadvantaged areas, which will address the knowledge gaps identified in the Nuffield-funded trial. The study will

also expand the applicability of the findings to a larger group of early years settings. Previous studies of *Talking Time*© have focused on the maintained sector only; this study will include early years settings from both the maintained and private, voluntary and independent (PVI) sectors. The results of the current evaluation will therefore inform educational initiatives undertaken in maintained settings as well as in PVIs.

The trial is supported financially and logistically by the Early Years Stronger Practice Hubs (SPH). SPHs were established in 2022 by the Department for Education (DfE) and are part of the early years education recovery support programme (EYER). Some hubs are led by an early years provider and others are led by a state-maintained setting. They support settings to adopt well-evidenced practice. They work with the EEF to select evidence-based programmes to fund and to make them available to settings. The following Stronger Practice Hubs are co-funding with EEF the delivery of *Talking Time*©: Liverpool City Region and Beyond Early Years SPH, A Brighter Start Early Years SPH, REACHout Early Years Stronger Practice Hub, and Thrive Together Early Years Stronger Practice Hub.

The *Liverpool City Region and Beyond Early Years SPH* aims to: establish local networks of early years educators to share early years knowledge and effective practice; act as a point of contact for early years curricular information; signpost to early years funded support programmes. It covers the local authorities of Cumberland, Westmorland and Furness, Halton, Knowsley, Lancashire, Liverpool, Sefton, St. Helens, Warrington, and Wirral. The *A Brighter Start Early Years SPH* supports settings to combine their professional judgement with robust evidence to improve outcomes for children and families. It covers the local authorities of: Barking and Dagenham, Barnet, Brent, Camden, City of London, Enfield, Hackney, Haringey, Harrow, Havering, Islington, Kensington and Chelsea, Newham, Redbridge, Tower Hamlets, Waltham Forest, and Westminster. The *REACHout Early Years Stronger Practice Hub* offers access to fully-funded training opportunities; access to experts to share strong, evidence-based practice; access to bespoke training; events and webinars on latest research in early years; visits to settings to enhance practice and to develop new approaches. It covers the local authorities of Bedford, Central Bedfordshire, Essex, Hertfordshire, Luton, Peterborough, Southend-on-Sea, Thurrock. The *Thrive Together Early Years Stronger Practice Hub* draws on a breadth and depth of early years knowledge and leadership including trusted experts and partners, SENCOs, PVI managers and childminders. The Hub's facilities include a bespoke training venue and access to a range of high-quality schools and group-based settings that exemplify best practice in early years' education and care. It covers the local of authorities of Dudley, Sandwell, Shropshire, Staffordshire, Stoke-on-Trent, Telford and Wrekin, Walsall, and Wolverhampton.[1]

This evaluation will assess whether the *Talking Time*© programme has an impact on the oral language development of children aged 3-4. The evaluation will comprise a randomised controlled trial, which will enable examination of the quantitative impact of the programme on children's language skills, and a process evaluation, to understand the extent to which the programme is implemented as intended, the perceived impacts on outcomes, and to add greater depth to the impact evaluation findings, including the extent to which there is support for the causal pathways set out in the underlying theory of change. Finally, the evaluation will

[1] See The Hubs | Stronger Practice Hubs.

also estimate the cost of the programme. Given the experimental nature of *Talking Time©,* a randomised trial is the ideal method to assess the effectiveness of the intervention in a robust way.

The study will generate more robust results than the previous randomised trial (Dockrell *et al.*, 2023) by evaluating the impact of the intervention in a context unaffected by major shocks, such as COVID-19, which affected the implementation and outcomes of the previous trial. The current study will produce results that can be more confidently attributed to the intervention and that will have higher external validity. In addition, the present study will rely on a larger sample of settings and children to increase the precision of the estimates; it will conduct a qualitative follow-up study six months after the end of the intervention to assess its long-term impact on educators' practices; and methods will be explored to assess the response of the outcomes to different children exposure to the quantity and quality of intervention delivery.

The protocol was registered with the ISRCTN registry with code ISRCTN84530173 (ISRCTN - ISRCTN84530173: Evaluation of Talking Time: efficacy trial)

Intervention

Intervention Name

Talking Time©

Why (theory and rationale)

Talking Time© is a universal oral language programme for children aged 3 to 5. It supports Early Years practitioners to deliver engaging, structured small-group activities to children and enhance oral language through high-quality interactions. It has three main elements: 1. A flexible intervention programme; 2. A practitioner manual, story books and resources; 3. Staff professional development (training + mentoring).

Children from more disadvantaged backgrounds on average perform more poorly in standardised assessments of language skills (Nelson et al. 2011, Law et al., 2018). This can have important consequences for the remainder of their education throughout the school system and for their academic outcomes beyond. While *Talking Time©* is a universal intervention (delivered to all children in the target age range within a setting), it is expected to particularly benefit disadvantaged children, given the typically higher prevalence of oral language needs among this group and thereby best described as universal-targeted. At the same time, there is scope to improve language-supporting practice among early years settings, including in settings that serve more disadvantaged populations. By providing high quality professional development to improve language-supporting practice, *Talking Time©* aims to improve the oral language skills of young children attending these settings. Due to its focus on building staff expertise, *Talking Time©* offers a sustainable approach to improving pupil progress.

Who (recipients)

The programme is a universal small-group intervention designed to benefit all 3–5-year-old children. The evaluation, however, will focus on children aged 3-4 at the start of the programme, to maximise the chance that the children will be attending the setting throughout

the delivery of the programme. Children younger than three at the time of collecting baseline data will not be included in the trial, although they may receive the intervention if they are in a room/class in which it is being delivered universally. The programme is delivered in both maintained and PVI nurseries but not in settings run by childminders.

The programme is universal at the child level but targeted at the setting level. It is universal at the child level because all eligible children in settings assigned to the intervention (either maintained or PVI) are included in the intervention. It is targeted at the setting level because the delivery team is focusing recruitment among settings that are likely to have a large proportion of disadvantaged children (although not all settings in the study will have a level of disadvantage above the national average). If the number of recruited settings exceeds the number of settings required for the trial, we will rank settings using the Income Deprivation Affecting Children Index (IDACI) from the English Indices of Deprivation 2019. We will then prioritise those settings that have a higher deprivation score.

What (materials and procedures)

Talking Time© consists of three components:

1. A flexible intervention programme for children.
2. The practitioner manual, story books and resources.
3. An embedded professional development programme for practitioners.

This integrated programme is designed to equip staff with the knowledge and skills needed for flexible, high-quality implementation using evidence-based ways of talking with children, and support them in adapting and embedding the programme into regular classroom practice.

The **programme for children** involves two 15-minute structured oral language sessions per week, for 20 weeks. Sessions are conducted in small groups (up to 5 children) of mixed language ability. There are three types of activity, all of which focus on supporting oral language through extended conversations with children (talking 'with' rather than 'to' children):

- *Story Conversations*: shared storytelling and conversation using the illustrations in storybooks as prompts;
- *Word Play*: games and guided role play designed to develop vocabulary breadth and depth;
- *Hexagons*: narrative discussion and retelling based on photos of real situations and routines likely to be familiar to children within their lives and local areas.

Two practitioners in each setting will be identified as *Talking Time*© leads and be responsible for delivering the majority of sessions.

The **professional learning component** supports the *Talking Time*© leads to:

- introduce and embed each of the three Talking Time activities
- use evidence-based language-supporting techniques during sessions to promote children's oral language
- tailor activities and practice to optimally meet children's needs for language support.

Overall, the aim is to gradually reduce reliance on the programme resources and materials, and support educators in adapting and embedding the programme and language-supporting interactions into regular classroom practice in the longer term.

The professional development package consists of:

- An initial welcome visit (which include contact with the setting leadership team)
- Three *twilight* training sessions, provided to all staff delivering the programme. These introduce practitioners to the *Talking Time*© activities, the underpinning framework of evidence-based language-supporting strategies and the professional development approach.
- Four in-class mentoring sessions for the *Talking Time*© leads, along with three online mentoring sessions for each team's pedagogical leader for *Talking Time*©. Mentoring (based on an 'instructional coaching' approach) supports staff to develop the knowledge and skills required to use *Talking Time*© prompts and plans flexibly, adapt them to the needs of specific children and groups, and use the language supporting strategies responsively and effectively to enhance children's oral language skills. The final in-person mentoring session and the online mentoring sessions support longer-term leadership, planning, and adaptation.
- Practitioners also engage in a weekly individual professional reflection; they are encouraged to complete one individual reflection each week and to engage in weekly *Talking Time*©-related professional collaboration; for example, shared planning, professional discussion or observation and analysis of sessions, using the "rehearsing, noticing, analysing and refining" reflective tool to refine their practice.

The professional development (PD) approach is designed to motivate; build knowledge; develop skills; and support embedding of language-supporting practice, based on the COM-B model of behavior change (Michie et al., 2011). It underpinned by a framework of six evidence-based language-learning strategies. Throughout the PD, the language strategy framework is used as a reflective tool to support educators in rehearsing, noticing, analysing and refining their practice.

Settings are also provided with a **manual that includes flexible plans and conversation prompts** for each activity, as well as ideas for wider classroom activities to reinforce learning. These are designed to provide prompts and cues help to scaffold sessions and use of the language-supporting strategies within them (see for example Sims et al., 2021). In addition, settings receive five picture books and a starter pack of hexagon photos to support the implementation of activities. In the final weeks of the programme, it is expected that practitioners will plan their own activities independently using their own choice of materials. In the longer term it is expected that they will go beyond the manualised intervention to embed language supporting activities and interactions into everyday practice in a manner which suits their setting and children.

The delivery mode may differ slightly in PVI settings, where the intervention is untested and the delivery of the above activities may require some adjustments. Such adjustments will become clear during the pre-trial phase in which the implementation of the programme will be field-tested in various PVI settings.

Who (provider)

The *Talking Time*© programme, including manual and activities, was originally developed by IOE at UCL (Dockrell, Stuart and King, 2010). Subsequently, a more comprehensive manual (updated and expanded to include core texts and flexible activity prompts and plans) and a comprehensive professional development programme were developed at the University of Oxford. These were integrated into the programme, in collaboration with IOE.

A team of trainer-mentors from the delivery team will provide the professional development and support participating early years settings to implement the programme. Each setting will be supported by a trainer-mentor who will conduct the initial orientation visits to build relationships and clarify expectations. Settings will designate two *Talking Time*© leads (expected to be a teacher and a teaching assistant or practitioner with equivalent role in PVI settings) who will be supported to lead the implementation and delivery of the sessions. The trainer-mentors will deliver twilight training sessions and provide in-class and online mentoring. Staff involved in the intervention will be encouraged to engage in self-reflection activities and will be kept in contact via weekly emails by the mentors. The setting-level activities will be delivered by the trained and supported staff. In large settings, additional staff, trained in the twilight sessions, may also lead the *Talking Time*© sessions. Alternatively, an additional TT Lead can be trained in larger settings.

How (format)

The professional development activities for educators are delivered in various ways. Initial visits by *Talking Time*© trainers and in-class mentoring are conducted in person. Twilight training sessions and the online mentoring sessions are provided online *Talking Time*© through video calls. Communications with leads and other staff are maintained via emails throughout the duration of the programme.

The *Talking Time*© activities for children are delivered to small, mixed-attainment groups of maximum 5 children. All children in a class are divided into groups of mixed language levels of around 5 children for the programme activities. The groups (maximum 5 children) are created by setting staff. The Talking Time handbook contains guidance for creating groups and trainer-mentors can provide further advice if required. The group sessions are interactive, and children are encouraged to engage with the various activities. The groups generally remain constant over the period of intervention although on occasion a setting decide to move a child and/or reformulate groups based on children's needs.

Where (location)

Children receive the *Talking Time*© programme in their early years settings during usual class time. The professional development for early years practitioners comprises both face-to-face and online elements. The face-to-face element will be carried out in the selected settings and classrooms.

This evaluation is co-funded by the Stronger Practice Hubs and will take place in five regions supported by the hubs: Thrive Together (West Midlands); Liverpool City Region and Beyond Early Years (North-West); REACHOut (East of England), A Brighter Start (East London), and St Edmund's Early Years Stronger Practice Hub (Yorkshire and the Humber). The St Edmund's Hub in Yorkshire and the Humber is not a funder of the programme, although some

local authorities fall within its area of operation. Settings located in areas under the St Edmund's Hub are funded by the Liverpool City Region and Beyond hub. The programme is implemented in some local authorities within the hubs. The full list of participating hubs and local authorities is included in Table 2 in the *Participant selection* section.

When and how much (dosage)

The *Talking Time*© programme lasts for 20 weeks and involves two 15-minute small group sessions for children each week.

The programme will be delivered to settings from late 2024 to Summer 2025. It will take place on a three-week rolling basis to assist with the practicalities of delivering the programme to settings (that is, the first third of settings will commence delivery in week 1, the second third will commence delivery in week 2, and the final third will commence delivery in week 3). In all cases the manualised programme is intended to run for 20 weeks (within school term times). Similarly, the end-line assessments will be conducted on a rolling basis to allow all children to be fully exposed to the programme.

Settings are encouraged to continue delivery beyond the scope of the 20 weeks. The professional development programme prepares them to do so but the delivery team do not provide further support beyond the 20 weeks. When sustaining the programme in future years (e.g. implementing with a new cohort of children), settings can choose to use the programme books again, or plan their TT activities with one/more of their own books e.g. those connected to their curriculum. Planning for sustaining Talking Time is the focus of discussions in the final in-person mentoring visit so Talking Time Leads feel prepared to implement independently.

Tailoring (adaptation)

A core feature of Talking Time© is the principle of flexible adaptation. Although manualised, the programme is designed to be flexible and adaptable to suit individual children, settings and contexts. The professional development is designed to support staff in tailoring the programme to meet the language needs of the children they are working with. The aim is to gradually decrease reliance on the provided programme materials, so that practitioners feel empowered to adapt the plans and prompts provided, plan activities of their own, and can progress to designing activities based on books and materials that they choose.

It is also worth noting that since the programme is untested in PVI settings, it is likely that some changes in delivery will be needed in these settings during and after implementation. It is difficult to anticipate what type of adjustments to the programme might be needed in PVI settings, but the changes will be studied and reported by the IPE component of the evaluation as described in the following sections.

Theory of change of the intervention

The logic model for the intervention is presented in Figure 1 below. Developed in the setup phase of the project, the logic model provides a visual representation of the theory of change underlying the programme. Inputs can be classified into two main groups; those that are targeted at the staff delivering the *Talking Time*© programme (forming the professional development package), and those that relate to the delivery of the programme to children.

Through the delivery of the professional development programme, it is expected that practitioners will engage with the various activities and resources, including the training manual, and the underlying strategies designed to improve their practice. These activities are intended to lead to improvements in language-supporting practice, as well as in practitioner confidence, knowledge and skills. This includes their confidence and ability in relation to the principle of flexible adaptation. Ultimately, the programme aims to improve language-supporting practice in the setting, resulting in successful implementation of the programme and in a sustained improvement in practice.

Through the delivery of the *Talking Time*© sessions, the programme aims to ensure children are exposed to language-supporting contexts and improved language-supporting practice, including increased opportunities for participating in meaningful conversations in small groups. These skills should also be reinforced in wider setting activities outside of the specific *Talking Time*© sessions. Through this greater exposure to improved language-supporting contexts and practice, children should experience improvements in vocabulary, grammar and narrative skills in the short-term, which ultimately will improve their overall literacy skills and academic attainment.

Since the programme includes a substantial professional development component – including elements of self-reflection and adaptation - it is expected that educators will tailor the activities of the intervention to the specific setting in which they operate. It is also expected that the programme activities, if successful, will be likely incorporated in routine practice and will therefore produce long-lasting benefits that outlive the implementation of *Talking Time©.*

We include two moderators which we believe are likely to significantly affect the impact of the intervention: children that have English as an Additional Language (EAL) and children eligible for the Early Year Pupil Premium (EYPP). The moderators will be used to conduct sub-group analyses.

Children with EAL are a heterogeneous group whose parents include economic migrants, refugees, highly skilled professionals, and high-income earners. However, the intersection for some of social disadvantage and of EAL status may create a double disadvantage. Attending school in a socially and economically deprived neighbourhood may affect language proficiency, and there is evidence that children learning English as a second language are at some risk of literacy difficulties (August and Shanahan, 2006; Kieffer, 2008). UK data suggest a relationship between educational attainment and English proficiency. For example, school census data show that only 43% of pupils recorded as being new to English achieved a good level of development (GLD) at the end of reception compared to 88% of those who were recorded as fluent.[2]

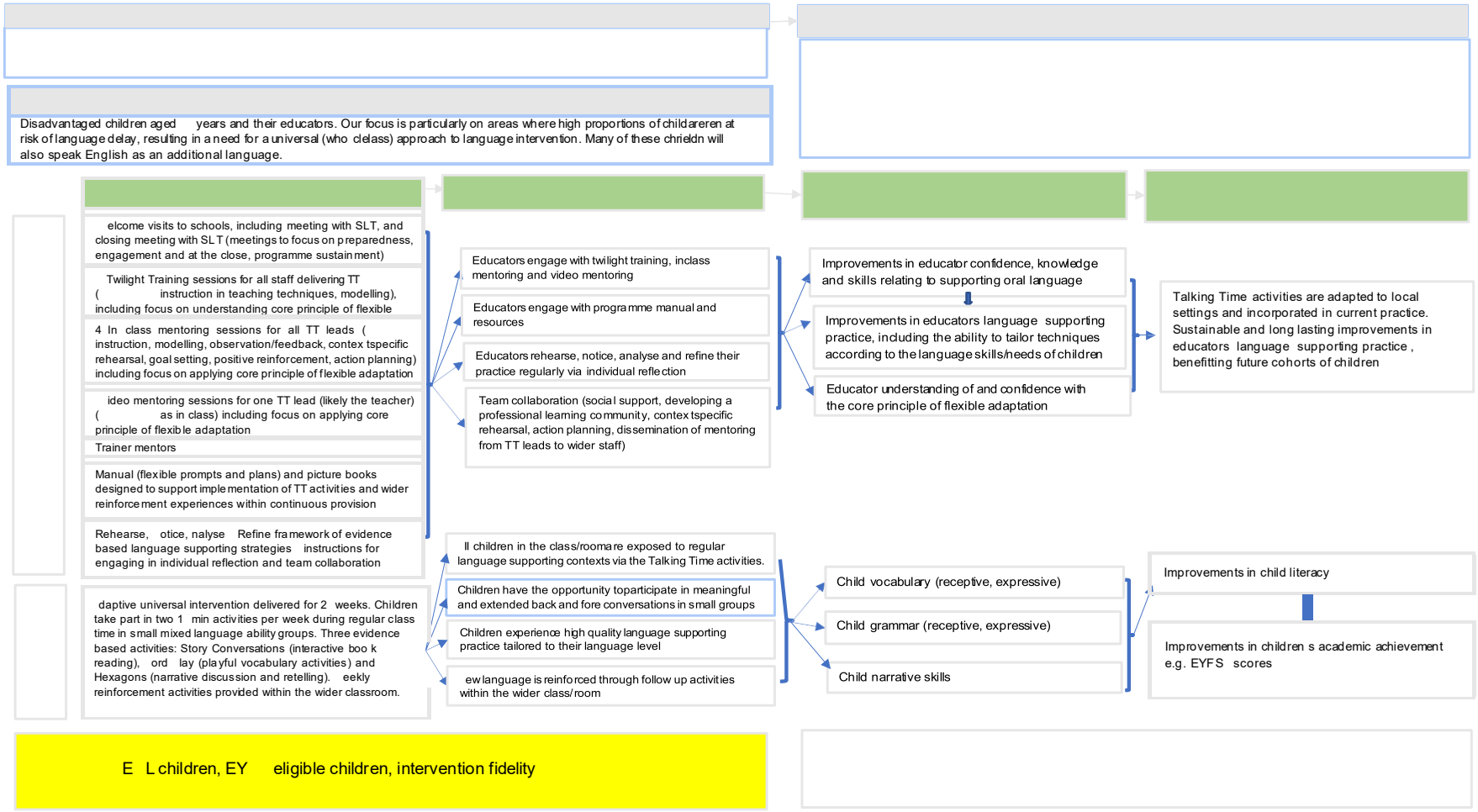
EYPP eligibility status is a broad indicator of low family income. There is evidence that children of low-income earners growing up in areas of social disadvantage experience disproportionate

[2] See DoF (2020), English proficiency of pupils with English as an additional language ((English proficiency of pupils with English as an additional language (publishing.service.gov.uk)).

educational delays relatively to their more advantaged peers (Nelson *et al.*, 2011; Law *et al.*, 2018).

The effectiveness of the intervention is also likely to be related with the level of compliance and with fidelity, i.e. the extent to which the intervention was implemented as planned. Children's attendance at the *Talking Time*© sessions is an indicator of compliance. However, the successful implementation of the intervention requires the active participation of settings' leads and staff, which in turn depends on the quality of the tools and methods of instructions and on the existing constraints at the setting level such as, for example, the presence of conflicting work demands, staff capacity, and the availability of basic and functional infrastructure. As discussed in the section on compliance, this will be measured in various ways. We will develop indicators of compliance and fidelity in order to assess to what extent they moderate the effectiveness of the intervention.

Figure 1: Logic model of Talking Time©



Impact evaluation design

Research questions

The primary research question that this impact evaluation is designed to address is:

RQ1. What is the impact of *Talking Time*© on children's oral language skills as measured by a composite index of expressive vocabulary (Renfrew Expressive Vocabulary Test) and of Information and grammar abilities (Renfrew Action Picture Test)?

The following sub-questions of this (RQ1) primary research question will also be explored:

> RQ1a. What is the impact of *Talking Time*© on oral language skills of disadvantaged children that are eligible for the Early Years Pupil Premium (EYPP)?
>
> RQ1b. What is the impact of *Talking Time*© on oral language skills of children with EAL?

The secondary research questions of this project are:

RQ2. What is the impact of *Talking Time*© on different aspects of children's oral language skills as measured by the subtests of Renfrew Expressive Vocabulary Test, Renfrew Action Picture Test and the sentence repetition assessment from the GAPS test?

The following sub-questions of this (RQ2) secondary research question will also be explored:

> RQ2a. What is the impact of *Talking Time*© on grammatical development, as measured by a sentence repetition assessment (GAPS), of children eligible for the Early Years Pupil Premium (EYPP)?
>
> RQ2b. What is the impact of *Talking Time* on grammatical development, as measured by a sentence repetition assessment (GAPS), of children with EAL?

RQ3. How does the impact of the intervention vary with compliance?

Design

Table 1: Trial design

Trial design, including number of arms			Two-arm, cluster randomised
Unit of randomisation			Setting
Stratification variables (if applicable)			Strata are five geographic areas covered by five Stronger Practice Hubs (SPH) and setting type (maintained versus PVI settings)
Primary outcome	**Variable**		Oral language skills
	Measure (instrument, source)	scale,	A composite index of Expressive vocabulary – Renfrew Expressive Vocabulary (RAV) and of Information and Grammar – Renfrew Action Picture Test (RAPT)
Secondary outcome(s)	**Variable(s)**		1. Sentence repetition 2. Expressive vocabulary 3. Information 4. Grammar
	Measure(s) (instrument, source)	scale,	1. Sentence repetition score component of the Grammar and Phonology Screening test (GAPS) scored from 0 to 11 (Gardner et al., 2006). 2. Renfrew Expressive Vocabulary (REV) scored from 0 to 100 (Renfrew, 2023). 3. Information – Renfrew Action Picture Test (RAPT) scored from 0 to 41, (Renfrew, 2019). 4. Grammar – Renfrew Action Picture Test (RAPT) scored from 0 to 39, (Renfrew, 2019)
Baseline primary outcome	**for** **Variable**		Oral language skills
	Measure (instrument, source)	scale,	A composite index of Expressive vocabulary – Renfrew Expressive Vocabulary (RAV) and of Information and Grammar – Renfrew Action Picture Test (RAPT)
Baseline secondary outcome	**for** **Variable**		1. Sentence repetition 2. Expressive vocabulary 3. Information 4. Grammar
	Measure (instrument, source)	scale,	1. Sentence repetition score component of the Grammar and Phonology Screening test (GAPS) scored from 0 to 11 (Gardner et al., 2006). 2. Renfrew Expressive Vocabulary (REV) scored from 0 to 100 (Renfrew, 2023). 3. Information – Renfrew Action Picture Test (RAPT) scored from 0 to 41, (Renfrew, 2019). 4. Grammar – Renfrew Action Picture Test (RAPT) scored from 0 to 39, (Renfrew, 2019)

This is an efficacy trial. The goal of the trial is to test how the intervention works under ideal conditions, not how it would work if implemented under actual conditions. The treatment is delivered in a standardised way and care is taken to ensure the fidelity of the intervention and that the treatment is delivered in full, in such a way that the chances of showing an impact are maximised.

The trial is designed as a two-arm, cluster-randomised trial. Settings randomised to the project group will receive the *Talking Time*© programme (first arm); settings randomised to the control group will not receive the programme and will be expected to continue with business-as-usual (second arm). Settings in the control group will receive an incentive payment of £1000 at the end or the trial. Although the cost of *Talking Time*© exceeds the amount of £1000 per setting, the delivery team will commit to providing control settings with a package of *Talking Time*© training/mentoring after the study is completed, which will offer substantial support and will be within the funds made available to them. All settings, regardless of whether they are allocated to the intervention or control group, will also receive a thank you payment of £400 for participating in the research activities (£150 in November 2024 for supporting the completion of the baseline assessments, and £250 in summer 2025 for supporting the completion of the endline assessments). In addition, the EEF will make a contribution towards covering time spent by setting staff attending training activities in the intervention settings at a rate of £7.50 per hour.

The intervention is designed as a setting-level intervention and will be delivered universally to all children in the relevant classes/classrooms or groups divided into small groups. In the case of some very large settings, it was not possible to deliver to the whole relevant year group (e.g. because it was not feasible to release all the staff for training at one time and/or because it was too challenging for settings to logistically plan delivery for large numbers of children whilst still getting to know the programme). For these settings a 'study class' or 'study room' was selected to participate, with the aim of rolling out to universal delivery in the longer term (i.e. after the time-frame of the evaluation) should the setting choose to continue with the programme. Where there were multiple classes/groups within the same cohort, one class or group was selected to participate and be randomly assigned to either the intervention or control group. Where a setting did not divide children into easily assignable classes/groups, an alternative structural feature was used to select a group for the intervention such as "key-worker groups" or all "morning children". A different type of design, such as an individual-level randomised trial, would have several disadvantages. First, if some groups were assigned to the intervention and other groups to the control group within the same setting, there would be a risk of contamination because the study could not prevent interactions among children belonging to different groups and school staff. Secondly, there would be considerable logistical difficulties if randomisation occurred at the group-level within settings. For example, only some members of staff would be trained, but not all; the intervention should be delivered at some specific times of the day or week or in a separate environment within the setting. Finally, the intervention is designed at the setting-level and the effect of the intervention on a single group of children within the setting under an alternative design might underestimate the true effect. This would be the case, for example, if training and coaching were more effective when delivered to all staff rather than to some staff only.

The trial aims to establish the impact of the *Talking Time*© programme on young children's oral language. The primary outcome is language skills (a composite index of Expressive vocabulary measured by the Renfrew Expressive Vocabulary Test (Renfrew, 2023) and of Information and grammar measured by the Renfrew Action Picture Test (Renfrew, 2019)).. The secondary outcomes are Expressive vocabulary, from the Renfrew Expressive Vocabulary test (Renfrew, 2023), Information, from the Renfrew Action Picture Test) (Renfrew, 2019), Grammar, from the Renfrew Action Picture Test) (Renfrew, 2019), Sentence repetition from the Grammar and Phonology Screening (GAPS) (Gardner et al., 2006) These measures

were chosen in collaboration with the delivery team, following discussion over the logic model and intended outcomes, with due consideration also given to practical factors such as minimising burden on participants. These outcome measures and criteria for selection are described in further detail later in this protocol.

Participant selection

The trial will take place in five geographic regions (East of England, London, North West, West Midlands, Yorkshire and the Humber) covered by five Early Years Stronger Practice Hubs (SPH) (Thrive Together; Liverpool City Region and Beyond; REACHOut and A Brighter Start, and St. Edmund's). Of these five SPHs, four are supporting the recruitment of settings for the programme and are co-funding the intervention (the exception is the St. Edmund's hub, whose financial contribution to the programme is provided by the Liverpool City Region and Beyond Hub). The hubs and the local authorities participating in the programme are listed in Table 2.

Table 2 Early Years Stronger Practice Hubs and local authorities participating in the programme

Hub	Local authorities
Thrive Together (West Midlands)	Dudley, Sandwell, Shropshire, Stoke-on-Trent, Telford and Wrekin, Walsall, and Wolverhampton
Liverpool City Region and Beyond Early Years (North-West)	Knowsley, Lancashire, Liverpool, Sefton
REACHOut (East of England)	Bedford and Central Bedfordshire, Essex, Hertfordshire, Luton, Southend, and Thurrock
A Brighter Start (East London)	Barking and Dagenham, Barnet, Brent, Camden, City of London, Enfield, Haringey, Harrow, Havering, Islington, Kengsinton and Chelsea, Newham, Redbridge, Tower Hamlets, and Waltham Forest
St Edmund's Early Years (Yorkshire and Humber)	Doncaster, Calderdale, Kirklees, and Wakefield

Settings will be recruited by the delivery team with the support of the SPHs and those responding with an expression of interest will be considered for inclusion. The intervention aims to target settings whose attending children live in disadvantaged areas and/or in low-income families. Settings expressing an interest to participate will be ranked using the IDACI deprivation index, and if the number of recruited settings exceeds the number of settings required for the study, the settings with higher values of the deprivation index will be selected for the programme.

Maintained nurseries and PVI settings are eligible for the programme. The following settings however are not eligible: settings participating in another EEF funded evaluation in the early years, settings participating in another funded SPH programme between 2022-25, and

settings run by childminders. All children in the selected settings and classrooms will be eligible for the intervention, with the exception of children attending the setting for fewer than 15 hours per week.

Recruitment

Recruitment is led by the delivery team at the University of Oxford, working together with the Stronger Practice Hubs, and with support from the evaluation team and the EEF. Recruitment documents were developed collaboratively by the delivery team, the evaluation team and the EEF. An initial setting information sheet provides settings with information about the project. Following discussions with the delivery team or attendance at an online information session, if a setting is interested in proceeding, and meets the eligibility criteria, the setting is asked to sign a memorandum of understanding which sets out what participation in the project entails and the roles of all parties.

Settings will be asked to distribute information sheets to parents of eligible children along with a privacy notice, and a consent form. Only children whose parents have signed and returned the consent form will be included in the study. The consent form will also cover parents' authorisation to record the language assessments. Audio-recording of assessment was introduced to allow assessors to re-listen children's answers and will improve the quality of the data. The form will also enable parents to provide consent for their child to participate in video recordings for the online mentoring element of the programme and audio recordings for an additional component of the study which investigates the impact of the intervention on staff practices and that will support staff to improve *Talking Time*© sessions.

To ensure the study yields results representative of the population, the proportion of settings selected in each region should align with the proportion of that region's population in the total population involved in the study. Recent UK census data indicates that the five regions considered in our study represent 60% of the total English population. Table 3 outlines the percentage of the study population in each region and the corresponding number of settings in each region to achieve representative results.

Table 3 Target number of study settings per region

Region	Proportion of study population	Population-based target settings	Actual recruitment target
London	0.26	34	14
North-West	0.22	28	30
East of England	0.19	24	43
West Midland	0.18	23	43
Yorkshire and the Humber	0.16	21	Included in the North West figure

Assuming a total of 130 settings for the study (as will be discussed in the section on statistical power), Table 3 suggests an ideal target number of settings per region based on population figures. The last column presents the recruitment targets based on SPH funding allocations.

Outcome measures

Baseline measures

All tests (RAV, RAPT, and GAPS) will be administered both at baseline and endline. Assessments will be carried out in one-to-one sessions with children by 40 assessors across the 130 settings. The assessors will be hired and trained by Qa Research from locally recruited university students. The assessors will score the assessments and Qa Research will monitor the quality of the scores provided.

The administration of the assessments will occur before randomisation takes place and assessors will be blind to intervention assignment. This will avoid that knowledge of assignment status affects the outcome of the assessments either in the intervention or in the control group. In fact, it is planned that assessors will be blind to the intervention assignment throughout the study, including at the time of endline assessments.

There are two reasons for administering the same tests before and after the intervention. First, the estimation of project effects using an ANCOVA model as discussed in the Statistical Analysis section (i.e. a model including pre-intervention assessments as a covariate) increases statistical power and therefore the chances of finding an impact if there is indeed an impact. Second, baseline values of the tests will allow the researchers to check how well balanced are the observations across the control and intervention groups. Information on differences in baseline tests scores is needed at the analysis stage to interpret the results correctly.

The assessments will be piloted in 5 settings in a pre-trial exercise in order to assess their acceptability and feasibility. The pilot will specifically investigate the following: the time required to complete the tests; the extent to which the assessments are capable of capturing children's variability in skills, and avoiding null scores; the most appropriate approach to code null assessments; and the suitability of the assessments in PVI settings.

Primary outcome

The primary outcome of the evaluation is oral language skills. Language competency is multidimensional, including elements of grammar, vocabulary, phonology, and narrative discourse (Massonnié *et al.*, 2022). The intervention is designed to improve children oral language skills along all these dimensions, except phonology. More specifically:

- *Story Conversations* involves joint story-telling and conversation with children using the pictures in storybooks as prompts. It is designed to promote comprehension, narrative and more complex language skills, such as describing, inference and predicting 'what next'. It also provides opportunities for learning new words.
- *Word Play* activities are designed to develop children's vocabulary depth and breadth through guided play in meaningful contexts;
- The *Hexagons* activity uses photos based on real life situations to give children opportunities to create and retell narratives and talk about their own lives. It is designed to build on and reinforce skills practised in Story Conversations, including conversation, narrative skills and talking about things outside of the here and now. It also offers opportunities to introduce and reinforce key words.

The three activities together are designed to promote vocabulary, comprehension, and oral narrative.

In order to capture the success of the programme along these dimensions we use a composite index of independent assessments of expressive vocabulary and expressive language skills. Expressive vocabulary will be assessed using the Renfrew Expressive Vocabulary Test (REV).[3] The REV test assesses to what extent a child can name pictures of words correctly in order of difficulty. . It is expected to take about 5 minutes to complete. Expressive language skills will be assessed using the Renfrew Action Picture Test (RAPT), which includes a grammar and an information assessment.[4] The RAPT assesses children grammar skills by asking them to describe a series of vignettes using words, verbs, and other grammar constructs of increasing difficulty (see a sample card in Appendix 1). The children's grammar outcome provides a proxy for quality of connected speech . The information score provides an additional measure of vocabulary knowledge.

Oral language is a general construct. It is multidimensional and it cannot be captured by any single assessment. No theory can guide the combination of different indicators representing the different dimensions of oral language, and we are not able to assign weights to the different components based on prior knowledge. We will therefore use principal component analysis to extract the principal component of the various indicator and we will use the extracted principal component to build an index of oral language skills. Principal component is a method developed precisely to uncover latent constructs, and it is particularly appropriate in this case in which no prior knowledge on the relevance of different components of oral language is available.

Secondary outcomes

The secondary outcomes of the study are 1) Expressive vocabulary, from the Renfrew Expressive Vocabulary test (Renfrew, 2023); 2) Information, from the Renfrew Action Picture Test) (Renfrew, 2019); Grammar, from the Renfrew Action Picture Test) (Renfrew, 2019); Sentence repetition from the Grammar and Phonology Screening (GAPS) (Gardner et al., 2006) The Renfrew assessments were described in the previous section. GAPS is an assessment devised for children from 3 years and 4 months to 6 and a half years old to evaluate their basic grammatical ability across sentences and word forms (see Appendix 1). The test should not be administered to children under the age of 3. It assesses skills that typically developing children should have fully acquired by the age of 3-4. The test is simple and can be administered by professionals and non-professionals. It normally takes 5-10 minutes to complete. The results of the test are better interpreted in terms of percentile scores rather than raw scores. The test is widely used in the literature and validation exercises have found the test highly accurate in identifying impaired versus non-impaired children up to age 6 (van der Lely, Payne and McClelland, 2011).

---

[3] https://www.routledge.com/Expressive-Vocabulary-Test/Renfrew/p/book/9781032162492. The REV test is commercial and cannot be included in the protocol.

[4] https://www.routledge.com/Action-Picture-Test/peechmark-Renfrew/p/book/9781138586208. See Appendix 1.

The sentence repetition component of the GAPS assessment consists of 11 questions. The child is invited to repeat a small set of sentences after the administrator's prompt and with the support of a visual aid booklet story. GAPS also includes a phonological component, in which the child is invited to repeat eight nonsense words. However, only the first component of GAPS (sentence repetition) will be included in the study, while the second part (repetition of nonsense words, which measure phonology) will not be included because *Talking Time*© does not target phonological skills. The GAPS assessment is available at: http://www.hvdl.org.uk/gaps/.

The GAPS test was selected as a secondary outcome because it provides a precise measure of grammar competence, which is one of the components of general oral language skills. GAPS has been validated as a good predictor of language difficulties (van der Lely, Payne and McClelland, 2011) and is widely used (benchmark values for the UK population are available in such a way that its results can be interpreted in a comparative way (Dockrell et al., 2023). Finally, the test is quick and easy to administer including by non-professional staff.

It is estimated that the three tests (REV, RAPT, and GAPS) should take between 20 to 30 minutes to complete. The time required to administer all the tests will be piloted in the pre-trial and compared to records from previous trials using the same tests.

The project will also try to assess to what extent the impact of the intervention varies with exposure. We will use record data of children's attendance of *Talking Time*© sessions to build a categorical indicator of low and high exposure. We will use this indicator to compare the impact of intervention between groups attending the session with different frequency.

Floor effects

The assessments described above may deliver many zero scores for various reasons. For example, children may not be able to take the assessment because they are unwell, or they may stop soon after starting because of the perceived difficulty. Such problems are more common among young children and zero scores are more likely to occur at the baseline than at the endline. They may lead to a truncated distribution of scores, where many scores are set to zero, or they may simply lead to many missing observations.

Floor effects can be addressed at the design stage and at the analysis stage. At the design stage the task is to select assessments, and a protocol for carrying out the assessments, that maximises non-zero results. In the pre-trial pilot, we will investigate the extent of floor effects for all tests and we will consider whether a different ordering of the test, or other aspects of the testing environment, such as the set-up and the timing, may affect the results.

At the analysis stage different strategies are possible. For example, if missing scores are generated at random, an estimation method for truncated distributions, such as the *tobit* model, can be used to estimate effect sizes. If scores are not missing at random but correlated with children characteristics that are correlated with the outcome, such as for example socio-economic background, then a two-step estimation model such as the Heckman selection model could be used. If many scores are missing at the baseline but fewer are missing at the endline, then a composite score of the assessments conducted at the baseline (or an imputed score) could be used as a predictor when estimating the impact of the intervention on the endline scores. Details of the different estimation methods to deal with floor effects will be discussed in the analysis plan.

To be able to estimate effects in the presence of floor effects at the analysis stage, however, we will need to collect data on the reasons for the missingness of the baseline scores. At the analysis stage it makes a difference whether a score is missing because of for example illness (a relatively random event) or because of a refusal to take the assessment (potentially correlated with socio-economic characteristics). A coding for missing results will be formulated after piloting the assessments in the pre-trial and it will be used when collecting the baseline data.

Sample size

Table 4 presents the MDES obtained from power calculations carried out using *Optimal Design 3.01.* In our calculations we made the following assumptions, in addition to the standard assumptions of a 2-sided statistical test at 5% significance level (alpha=0.05) and of 80% power (1-beta=0.20 is the Type II error):

- A pre-post test correlation of 0.40. EEF research (Singh *et al.*, 2023) found the correlation between pre and post-test English tests between Early Years Foundation Stage and Key Stage 1 (the closest comparison to our assessments) in the range 0.43-0.63 using the NPD data, and in the range 0.54-0.88 using the data from past EEF-funded trials. We use the most conservative of the estimates in these ranges.
- An intra-cluster correlation coefficient of 0.10. EEF research suggests an ICC in early years trials at around 0.06 for studies where English is the outcome measure (Singh *et al.*, 2023). We adopt a more conservative value of 0.10 for our calculations.
- We also assume that the blocking variables account for 10% of group variance. Blocking in our study is obtained through stratification. We have no evidence to back up the assumption of 10% of explained variance. However, the assumed 10% effect of the blocking variable is in line with similar assumptions made in other studies, and we are not using the blocking variable in estimating MDES in the main results presented in Table 4. The impact of the blocking variable in Table 4 is therefore effectively assumed to be zero.
- The intervention is designed to reach 65 settings and the expected overall number of settings, including the control settings, is 130.

Figure 2 shows the MDES as a function of children per cluster, holding the number of clusters fixed at 130. The curve enters a relatively flat region around 10-12 children, meaning that there is limited benefit in increasing the number of children per cluster beyond 10 and that, conversely, less than 10 children per cluster decrease statistical power considerably.

Figure 2 MDES by number of children per setting

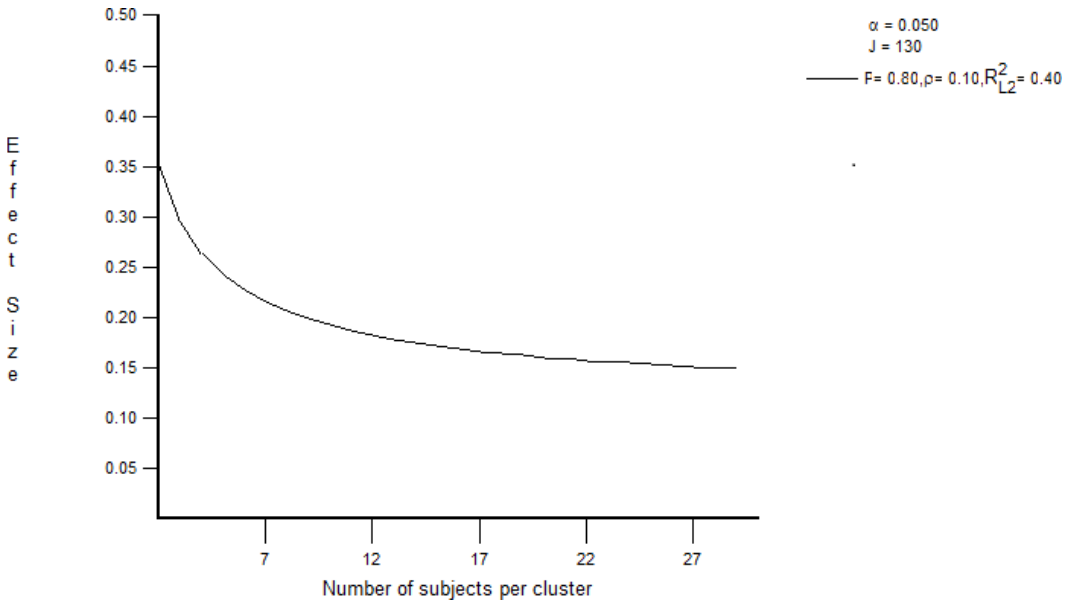


Table 4 presents the estimated MDES under three different scenarios regarding the number of children assessed. The number of children assessed in each setting may vary depending on the size of attrition (children being assessed either at baseline or endline but not in both) and the size of withdrawal (parents deciding not to take part in the study).

- Scenario 1. There is an average of 20 children per setting. In the Nuffield trial of *Talking Time*© the average class size was 28. Considering that some PVIs may have small group sizes, this is a reasonable estimate for a best-case scenario of small attrition and small withdrawal rates.
- Scenario 2. An intermediate case of moderate attrition and moderate withdrawal rate, which are together equal to 25%. In this case there is an average of 15 children per setting.
- Scenario 3. A combined attrition and withdrawal rate of 50%. As for comparison, this is the rate observed in the Nuffield-funded trial of *Talking Time*©, in which only 14 of the average 28 children per setting were included in the study. The Nuffield trial was adversely affected by the COVID-19 pandemic in addition to a high withdrawal rate due to the difficulty in obtaining parents' opt-in consent to participate in the study.

Table 4 shows that even a sample of 10 children per setting delivers an acceptable MDES smaller than 0.2 standard deviations. It should also be noted that these estimates were obtained under some conservative assumptions (no effect of the blocking variable, a small pre-post assessment correlation, and a relatively large ICC), and that on average statistical power in our sample will be increased by improving the quality of the data. Assessments will be recorded, and the recordings will be used to verify the answers and to monitor the quality of the assessors' scoring. Better quality data increase power in two ways. First, a lower measurement error implies a smaller sample standard deviation and therefore a larger detectable effect size (which is expressed in terms of standard deviation units). Second, a lower measurement error implies a stronger correlation between pre- and post-intervention assessments, which also increases statistical power.

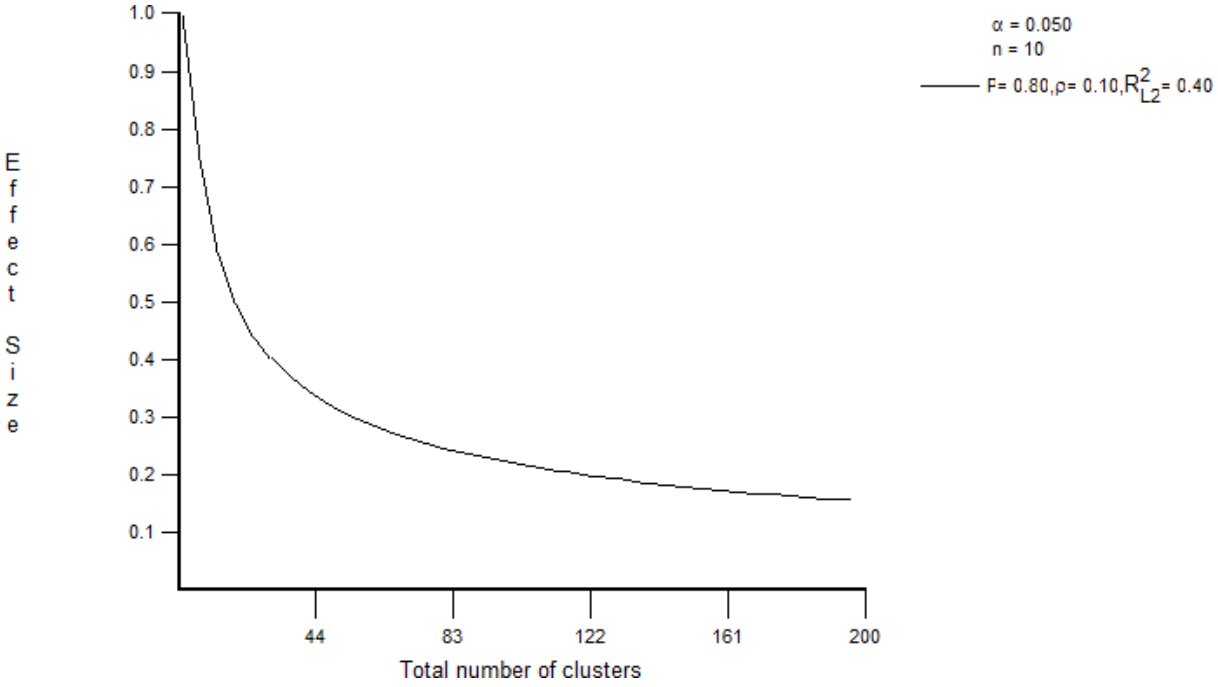
Table 4: Sample size calculations

			Best-case (n=20)	Intermediate (n=15)	Worst-case (n=10)	EYPP (n=3)
Minimum Detectable Effect Size (MDES)			0.16	0.17	0.19	0.19
Pre-test/ post-test correlations		level 1 (pupil)	0.4	0.4	0.4	
		level 2 (class)				
		level 3 (setting)				
Intracluster correlations (ICCs)		level 2 (class)				
		level 3 (setting)	0.1	0.1	0.1	
Alpha			0.05	0.05	0.05	0.05
Power			0.8	0.8	0.8	0.8
One-sided or two-sided?			Two-sided	Two-sided	Two-sided	Two-sided
Average cluster size			20	15	10	3
Number of settings		Intervention	65	65	65	65
		Control	65	65	65	65
		Total	130	130	130	130
Number of children		Intervention	1300	975	650	195
		Control	1300	975	650	195
		Total	2600	1950	1300	390

Our data collection plan is based on the expectation that the intermediate scenario will occur, in which all settings will be surveyed and at least 15 children will be assessed after accounting for moderate withdrawal and moderate attrition. This scenario will deliver an MDES of 0.17 standard deviations. However, we show that even under some worst-case circumstances (large attrition/withdrawal and moderate setting-level attrition), the study will be able to detect a reasonable effect size of about 0.2 standard deviations.

Note that even a reduction in the number of settings will allow the estimation of a reasonable effect size. Figure 3 shows the MDES in the worst-case scenario (of 10 children per cluster) against the number of settings surveyed. A 10% setting level drop-out rate (equivalent to a total of 13 settings) will result in an MDES of 0.20, a 20% drop out rate (26 settings) will result in an MDES of 0.22, a 30% drop out rate (39 settings) in an MDES of 0.23. As for comparison, the Nuffield-funded trial suffered a setting-level drop-out rate of 10% (4 schools withdrew from the study out of 40 originally recruited).

Figure 3 MDES by number of settings



The fourth column of Table 4 presents the MDES for the subgroup of children eligible for the Early Years Pupil Premium (EYPP). Given the age group of the children, this measure is used instead of eligibility for free school meals, in line with EEF guidance. Department for Education statistics indicate that in 2023, 14 per cent of children aged 3-4 old not in reception were in receipt of EYPP (DfE, 2023).[5]. Given the intention of the trial to target disadvantaged areas, our MDES calculations assume 20 percent of children will be eligible for EYPP; this is equivalent to an average of three children per setting using the intermediate attrition scenario of column 2. Keeping all other assumptions the same as in our overall scenario (column 1), this results in an MDES for this subgroup of 0.29. As an efficacy trial, the trial is not sufficinently powered to detect an effect on the EYPP subgroup.

Randomisation

Randomisation will take place at setting level. Settings will be randomised so that approximately half of settings will be assigned to the project group and half to the control group (who continue with "business-as-usual" and will receive an incentive payment for participation in the trial).

Settings will be randomised within strata defined on the basis of geographic area (with an average of two areas within each of the five Stronger Practice Hubs) and nursery setting type (maintained versus PVI). This stratification is used to reduce the risk of imbalance between the project and control groups.

Randomisation of settings, to achieve a 50:50 allocation, will be performed as follows:

[5] Available at: https://explore-education-statistics.service.gov.uk/find-statistics/education-provision-children-under-5#releaseHeadlines-tables

- Each setting will be assigned a randomly generated number
- Settings will be sorted by strata and random number
- The first setting will be allocated to either project or control, on the basis of a randomly-generated number
- Each subsequent setting will be assigned to have the opposite allocation of the previous setting.

The randomisation will be conducted by the evaluation team, using Stata 17. The randomisation process will be recorded in the syntax and log files used to carry out the randomisation. Appendix 2 includes a simple Stata code to illustrate how settings can be randomised within strata.

The results of the randomisation will be shared with the delivery team so that the programme can be provided to those settings allocated to the project group. This assumes that baseline assessments will be completed by the end of the Autumn half-term, with the exception of some mop-up visits. In the case of delays in obtaining the results of the assessments, we will consider randomising before the completion of all assessments thus allowing a small proportion of endline assessments to be conducted after randomisation. This presents a relatively small risk. The main reason for randomising after the assessments is blinding and both schools and assessors will be blind to randomisation for the whole duration of the assessment period. Analysts, however, will not be blind to trial arm allocation.

Statistical analysis

Primary analysis

The primary analysis will estimate the impact of the programme on oral language skills using a composite index of expressive vocabulary (Renfrew Expressive Vocabulary Test) and of Information and grammar abilities (Renfrew Action Picture Test)(RQ.1). The primary analysis will also analyse the impact of the intervention on oral language skills among EYPP disadvantaged children (RQ1a) and among children with EAL (RQ1b).

The estimated impact will be based on the difference between those assigned to the project and control groups, regardless of contamination of the control group or drop out. This is in order to estimate the “intention-to-treat” (ITT) effect. We will estimate outcomes using a linear regression model including a dummy variable indicating trial arm allocation and clustered standard errors at the setting level. We will use an ANCOVA (analysis of covariance) model, whereby the post assessment score will be regressed against the pre-intervention assessment score. The inclusion of the pre-test assessment (preferably in the form of raw score rather than its standardised correspondent value) in the model as an explanatory variable will increase the precision of the estimates and the statistical power of the analysis. The regression model will also include dummy variables for all strata, of which there will be 10 resulting from the product of five regions times the PVI status of the setting. As a robustness check we will also run a multi-level regression model to account for the hierarchical structure of the data (children within settings within regions). Further details of the statistical analysis will be provided in the Statistical Analysis Plan (SAP) after randomisation is complete. Analysts will not be blind to group allocation.

Secondary analysis

The secondary analysis will estimate the impact of the intervention on 1) Expressive vocabulary, from the Renfrew Expressive Vocabulary test (Renfrew, 2023); 2) Information, from the Renfrew Action Picture Test) (Renfrew, 2019); 3) Grammar, from the Renfrew Action Picture Test) (Renfrew, 2019); and Sentence repetition from the Grammar and Phonology Screening (GAPS) (Gardner et al., 2006).The secondary analysis will also analyse the impact of the intervention on all outcomes above among EYPP of disadvantaged children (RQ2a) and of children with EAL (RQ2b). The statistical approach adopted in the secondary analysis will be the same adopted in the primary analysis.

Estimation of effect sizes

The impact estimates obtained using the ANCOVA regression model will be converted into standardised effect sizes in order to facilitate the comparability of effects across studies. The effect size will be standardised by the unconditional pooled standard deviation of the project and control groups as recommended by the EEF statistical guidance (EEF, 2022b):

$$ES = \frac{(\bar{Y}_t - \bar{Y}_c)_{adjusted}}{sd_{unconditional}}$$

where $(\bar{Y}_t - \bar{Y}_c)_{adjusted}$ denotes the ANCOVA difference in means between the treatment and the control group adjusted for pre-test score and other stratification variables specified in the regression model described in the previous section. The pooled unconditional standard deviation is:

$$S_{unconditional} = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

In the case of GAPS we might be able to standardise the effect size by a population-level measure of the standard deviation rather than by the pooled unconditional calculated using the sample data.

The effect sizes estimated using multi-level modelling in the robustness check will be standardised in a different way. We will use a standard deviation based on the total variance to account for the nested structure of the data and for potential differences between settings, so as to avoid an overestimation of the effect size (EEF, 2022b).

For completeness, we will also report effect sizes as *Hedges g*, by adjusting the standardised effect size by the factor:

$$J = \frac{3}{4df - 1}$$

where *df* are the degrees of freedom used in the calculation of the unconditional pooled standard deviation above (Borenstein *et al.*, 2009). In addition, we will provide in the report all the parameters used in the calculations (sample size, variances, and ICC in each group) so that readers will be able to compute alternative definitions of effect sizes. The ICCs will be

calculated at the setting level of clustering for the post-assessment and pre-assessments using the unconditional variance.

Ninety-five per cent confidence intervals will be reported around the resulting effect size, calculated by inputting the lower and upper confidence limits for the coefficient on the treatment variable from the regression model into the effect size formula.

Sub-group analyses

We will conduct two subgroup analyses: by EYYP eligibility (RQ1a and RQ2a) and by EAL status (RQ2a and RQ2b). Given the age of the children, this study will analyse data for children who are eligible for the Early Years Pupil Premium, instead of the FSM subgroup which is typically used in EEF evaluations involving school-age children. Data on EYYP status will be collected from settings.

We will also conduct a subgroup analysis for children with EAL. As discussed in the Theory of Change section, children with EAL in the selected settings may face a double disadvantage, being both from impoverished families and being at higher risk of developing language difficulties. Information on EAL status cannot be obtained from the NPD datasets and was considered too sensitive to be collected from parents. Instead, it will be obtained from the settings, which routinely collect this type of information for census purposes.

In line with EEF Statistical Analysis Guidance, two approaches will be used and reported for the subgroup analysis. Firstly, a subgroup analysis will be conducted through estimating a separate model for children eligible for EYPP/EAL (that is, replicating the primary analysis model specification for EYPP/EAL children only). Secondly, we will repeat the primary analysis specification but additionally include an indicator for EYPP eligibility/EAL and an interaction term for EYPP eligibility/EAL and treatment status.

In order to avoid the cherry-picking of statistically significant result and statistically significant effects occurring as the result of chance, we will apply to the set of estimates in the sub-group analysis the correction for multiple testing of hypotheses described above.

Analysis in the presence of non-compliance

Data to assess compliance will be collected by the delivery team as part of implementation records. Compliance will be defined at the child level by identifying “treated” children in the intervention arm. In addition, we will develop indicators of fidelity of the intervention at the setting level based on data on practitioners’ attendance of professional development courses. A final subgroup analysis will be conducted using an indicator of intervention fidelity (RQ3).

The child-level compliance measure will be based on the number of sessions attended by each child. We will define the minimum number of sessions needed to produce some impact, and the number of sessions required to produce a reasonably good impact and we will use this information to build a measure of increasing compliance. The details of this measure will be discussed in the SAP.

We will use the child-level compliance indicator to conduct a Complier Average Causal Effect (CACE) analysis of the impact of the intervention on the treated children. This will be conducted for the primary outcome only. An instrumental variable (IV) approach will be used.

We will estimate the CACE using two-stage least squares (2SLS) regression by estimating a (first stage) model of compliance, using the measure of compliance described above and group allocation as the IV. The predicted values from the first stage are then used in the estimation of a model of our outcome measure. As per EEF analysis guidance, results for the first stage will be reported alongside with the correlation between the instrument and the endogenous variable and results of a F test. This analysis will be conducted using the *ivregress* command in Stata 17, which allows the necessary adjustments to be made to standard errors in response to the clustered nature of the data.

Non-compliance and the fidelity of the intervention will also be explored as part of the implementation and process evaluation.

Additional analyses and robustness checks

The delivery team are exploring the possibility of obtaining additional funding to collect audio recordings of *Talking Time*© sessions to explore potential impacts of the programme on staff practice. This analysis would be conducted by the delivery team. If it proves possible to obtain this additional funding, it may be valuable to explore whether these data can be used in combination with the data collected on children's language outcomes as part of this evaluation. This would enable exploration of the causal pathway set out within the logic model, exploring whether changes in practice mediate changes in children's outcomes.

Missing data analysis

Missingness in the data is most likely to arise through non-completion of the baseline assessments. It will be important to consider and document both the extent and pattern of missingness.

First, we will calculate and report the number of complete observations, and we will illustrate how the observations were lost using an illustrative flow diagram. Second, we will compare missing observations between the project and the control group to detect the presence of any differential attrition that could bias the sample. Third, when the proportion of missing observations is large, we will explore the main factors behind it.

More specifically, in the event of greater than 5% missing data at either cluster or individual level we will conduct further investigation into the mechanisms of missingness (EEF, 2022b). If there is less than 5% missing data, then it is unlikely that missing data will result in bias in the impact estimates (Schafer, 1999). Where more than 5% of data is missing, we will explore whether missingness is associated with characteristics observed at baseline, for example, whether those with lower assessment scores at baseline were more likely to be missing from the final sample.

First, we will assess whether missing data are MCAR, that is completely randomly generated. This will be obtained using a logistic regression of an indicator variable for missingness, on a series of covariate characteristics of the observations, such as age, sex, or socio-economic status. If the data are MCAR there will be no statistically significant coefficients estimates.

Second, we will assess whether missing data are MAR, that is missingness is independent of assignment to the intervention or the project group. This will be obtained by testing the difference in the proportion of missing data in the project and in the control group, and by the

t-test of the coefficient for treatment exposure in the model used to assess MCAR above. A no significant t-statistic in both cases will suggest that missing observations are unrelated to the intervention assignment.

Third, we will assess whether missingness is MNAR, that is not randomly generated across groups. This analysis will be performed using logistic regression model separately run for the intervention and for the control observations. A likelihood ratio test will be used to test the difference between the coefficients in the two models and to test that the two models are structurally similar. A low likelihood ratio test of the difference between the two models will suggest that the determinants of missing data are not different in the intervention and control groups.

A particular type of missingness may arise through floor effects if many children, for various reasons, will not be able to take part or to complete the assessments. As mentioned in an earlier section, we will employ statistical models to estimate intervention effects in the presence of floor effects, such as models for truncated distribution like the *tobit* model (if missingness is exogenous) or two-step models such as the Heckman selection model (if missingness is endogenous, that is if it is MNAR as defined above).

The approach to missing data analysis will be specified in more detail in the SAP. We anticipate that any such analysis will be focused on the primary outcome, in line with EEF analysis guidance.

Implementation and process evaluation (IPE) design

The aim of the IPE is to contextualise the findings of the IE and to explain how and why those findings have occurred. It will establish fidelity in implementation, identify the factors influencing impact and contextualise the impact evaluation. It will do this through mixed methods, using a survey and interviews with practitioners, trainer mentors, and senior leaders at the settings to establish their experiences of and perspectives on *Talking Time*© in terms of its content, their delivery of it, and the children's responses. Such a focus will interrogate the causal assumptions of the logic model (see Figure 1 above). We will thereby explore programme effectiveness, perceived impact of different elements of the intervention and considerations for a wider roll-out.

The IPE and the IE will be independently conducted by different members of the evaluation team with relevant expertise. However, the results of the IPE and of the IE will be integrated as follows: the evaluation team will hold internal workshops after the pre-trial, after the baseline assessment, and after the production of the first results following the collection of the endline data. During the workshops, members of the team working on the IE and IPE will share their findings and define the strategic direction of further work. In particular, they will identify areas of agreement, areas of disagreement (and following reconciliation of findings), and areas which deserve further investigation. The IPE may inform the IE in relation to the exploratory analysis of the impact of the interventions on sub-groups or on the actions of specific impact mediators (such as for example, the degree of fidelity or a specific project activity, whose details will be defined in the SAP) and by probing the quantitative findings in the final follow-up study which occurs six months after the completion of the intervention (such as for example, whether some changes in staff behaviours are sustained).

Research questions

We present the IPE research questions under the broad dimensions of fidelity, context and moderators, and perceived impact. As this is an efficacy trial, the main focus of the IPE is on fidelity. We need to know whether the intervention was implemented as planned, and whether there were significant deviations from implementation plans. A secondary goal of the IPE is understanding the validity of the logic model postulated, and learning lessons for a potential future effectiveness trial. Therefore, we also explore the contextual dimension to better understand which factors affected the success of the intervention. An important factor to the success of the intervention are practitioners' and senior leaders' attitudes and expectations, we therefore include the investigation of a further dimension of perceived impact. Furthermore, to understand the particular contribution of *Talking Time*©, we also include questions on programme differentiation – that is, the extent to which *Talking Time*© differs from both 'business as usual' and other literacy interventions that settings may have hosted.

Fidelity/Adherence

To establish fidelity in implementation, we will explore whether the programme is delivered as intended and examine what fidelity and adherence to the intervention consists of for the different parties, including the *Talking Time*© mentors, and how well settings have achieved this. We will also investigate the sub-dimension of usual practice in order to understand the status of the business-as-usual case, and whether the intervention produced any unexpected effects either in the intervention group or in the control group.

The IPE will also compile data on compliance, which will be instrumental to a) conclusively defining compliance and to estimating the CACE effect, b) defining levels of compliance that are associated with expected outcomes (for example the minimum amount of sessions likely to generate an impact and the number of sessions likely to generate a reasonable impact) to be used in the dose-response analysis of programme impact.

IPE RQ 1: To what extent were the programme's core components implemented as intended? Elements to explore include:

- Content (i.e., what is covered in the sessions?)
- Frequency (i.e. how often were the sessions delivered? Was the frequency of the three main activities as intended, i.e. did it match the programme timetable?)
- Duration (i.e. did settings deliver sessions for longer/shorter than specified? Were sessions delivered over the intended number of weeks)
- Dosage and compliance among children: how many sessions were attended by each child for each of the three session types in the programme settings? Were sessions delivered in small groups as a core component of the programme? Did practitioners deliver universally to all children in the 'study group'?
- Dosage and compliance among practitioners: how many professional development sessions were attended by practitioners in the programme settings? Which ones?
- Adaptation (i.e. do practitioners adapt any aspects of the programme? What are the reasons for the changes? How congruent with programme principles are these adaptations?). The programme allows for some degree of flexibility in implementation and is intended to be self-sustaining after support is withdrawn. So, for example,

practitioners are encouraged to adapt plans/prompts so sessions are tailored to the needs of the children. However, some adaptations (e.g. concerning timetabling and/or sequencing of activities, universal delivery, majority of delivery by the lead practitioners, and/or practitioners' engagement with the CPD) may not be entirely congruent with the programme principles. It is important therefore that adaptations, and rationales for them, are identified and documented.

IPE RQ 2: Is the intervention feasible for practitioners to engage in and implement as intended? Does this vary by setting type (i.e. nursery school/class or PVI)?

IPE RQ 3: How do practitioners engage with the training and the follow-up mentoring support? Which practitioners are selected to attend training and deliver the programme, and how? Is *Talking Time*© content accessible to all early years practitioners from different types of early years providers? What kinds of training and support do they value from trainer-mentors? Do all practitioners who deliver *Talking Time*© sessions attend the training, and whether/how do *Talking Time*© lead practitioners support other colleagues who may not directly access training?

IPE RQ 4: Do settings plan to retain *Talking Time*© after the intervention is complete? If there are plans for the continued delivery of *Talking Time*© in the setting, what adaptations, if any, would take place there to sustain *Talking Time*© within routine practice? How confident do practitioners feel about the possibility of delivering the programme without external support?

Context and moderators

We will consider the wide range of issues which affect implementation. Implementation research questions will consider broader dimensions than those under Fidelity above, including the identification of enablers and challengers in specific settings that contribute to successful implementation, as understood by practitioners, senior leaders and trainer-mentors.

IPE RQ 5: What are the potential barriers and facilitators to the delivery of the programme? Do these vary by setting type?

IPE RQ 6: How are the different elements of the programme (training sessions, in-class mentoring sessions, resources, etc.) perceived and experienced by early years practitioners? Are there relevant differences in perception across different types of early years providers?

IPE RQ 7: What contextual factors support successful take-up of *Talking Time*©? (e.g. is there enough support provided by setting leaders and trainer mentors to practitioners to implement the programme as intended? To what extent do practitioners feel ready for and open to the programme? Are practitioners able to develop successful team working practices to deliver the programme and, if not, what are the barriers to this?)

IPE RQ 8: Are there any specific facilitators/barriers for different groups of children, (e.g., children from disadvantaged backgrounds/EAL) to access and engage in *Talking Time*© activities?

Perceived impact

e will explore practitioners' (including setting leaders') expectations and perceptions around outcomes. It will be important to explore how practitioners perceive the impact of the programme on teaching strategies and children's outcomes, including any lasting impacts and whether and how practitioners change their practices or plan activities independently. We will explore the following potential outcomes identified by the Theory of Change:

IPE RQ 9: Do practitioners perceive that children are engaging positively with the sessions? Do practitioners perceive any variation in engagement by children's characteristics, e.g. EAL, additional learning needs?

IPE RQ 10: Is there evidence among practitioners of increased understanding, knowledge, skills and/or confidence to support children's early language development? Does practice change as a result, and in what ways?

IPE RQ 11: Are practitioners using strategies learned and/or programme activities to support children's language and communication during the intervention? Do they use these strategies/activities outside of the *Talking Time*©?

IPE RQ 12: hat are practitioners' perceptions of any outcomes and impacts on children's language and communications skills? Does the level of perceived impact differ for different groups, e.g., children from disadvantaged backgrounds - EYPP eligibility – and /or EAL?

IPE RQ13: Are there any unintended consequences or negative effects of the implementation of *Talking Time*©, relating to attendance, staff turnover, parental engagement, children's behaviour, or other factors?

Programme differentiation

IPE RQ 14: What was usual practice in relation to oral language development in all settings (including control settings) prior to the intervention? To what extent did 'business-as-usual' differ between nursery school/class and PVI settings? Are there any changes in the practice of control settings during the trial? What, if any, other language intervention programmes had settings delivered? How did *Talking Time*© compare to 'usual practice' in terms of accessibility (for practitioners and children) and effectiveness in terms of perceived impact?

These research questions will add rich and nuanced detail to the IE findings. The IE will establish compliance with the programme and its effect on the children. The IPE will answer some of the *why* questions behind these results. It will identify the degree of fidelity with which the programme was implemented in different settings, and also seek to answer why that degree of fidelity was achieved and what variables impacted this. The IE will determine the effectiveness of the *Talking Time*© intervention in terms of the children's development, and the IPE will explain why it had that degree of effect in different settings, focusing on contextual factors including the division of settings between nursery settings/classes and PVI, and practitioner perceptions of and responses to the intervention.

Research methods

Drawing on the EEF IPE guidance (EEF, 2022a), we propose an evaluation design based around a triangulation of qualitative and quantitative methods, to examine the research questions above (see Table 5). To do this we will conduct:

- A pre and post intervention survey to address particularly RQ1 on compliance and RQ 14 on programme differentiation
- 30 in-depth interviews with early years practitioners (teachers, teaching assistants or early years practitioners) (to address Fidelity, Implementation and programme differentiation). We will interview only one *Talking Time*© lead per setting, although, when possible, we may conduct paired interviews, because the pilot study found instances where practitioner interactions in interviews, produced valuable data. A total of 30 interviews is chosen as a manageable sample in the timeframe and a reasonable percentage of the whole, balancing representativeness and saturation. We do not expect saturation to be an issue, given the variation between settings, but if saturation is achieved before interviewing 30 practitioners, the interviews will stop.
- In-depth interviews with all the trainer-mentors (expected to be 7/8) from the delivery team (to address Fidelity, Implementation RQs and gather trainer-mentor perspectives on Perceived Impact)
- 20 in-depth interviews with senior leaders in participating programme settings, (to address Fidelity, Implementation and Perceived Impact)
- 10 in-depth interviews with setting leaders in control group (to address Programme Differentiation, Monitoring and Context/Moderators)

This distribution of interviews, we believe, focuses on collecting the detailed practitioner experiences of delivering the programme, whilst including the perspectives of setting leaders (in programme and control groups) and trainer/mentors.

Table 5 IPE methods overview

IPE dimension	IPE RQ addressed	Research methods	Data collection methods	Sample size and sampling criteria	Data analysis methods
Fidelity/Adherence	1, 2, 3, 4, 6, 14	Practitioner survey	Online pre and post intervention survey	65 settings	Survey analysis
		Interviews	Semi-structured interviews with practitioners, setting leaders, and trainer mentors	20 setting leaders (project group) 30 practitioners 7/8 trainer mentors	Framework analysis
Adaptation	1,2, 6, 7, 14	Practitioner survey	Online pre and post	65 settings	Survey analysis

			intervention survey		
		Interviews	Semi-structured interviews with practitioners, setting leaders, and trainer mentors	20 setting leaders (project group) 30 practitioners (project group) 7/8 trainer mentors	Framework analysis
Dosage	1,2,3,4	Practitioner survey	Online pre and post intervention survey	65 settings	Survey analysis
		Interviews	Semi-structured interviews with practitioners, setting leaders, and trainer mentors	20 setting leaders (project group) 30 practitioners 7/8 trainer mentors	Framework analysis
Quality	3, 5, 6, 10, 13	Interviews	Semi-structured interviews with practitioners, setting leaders, and trainer mentors	20 setting leaders (project group group) 30 practitioners (project group_ 7/8 trainer mentors. These interviews are particularly relevant to Quality as they have experience of TT in more than one setting	Framework analysis
Reach	4, 7, 8, 13, 14	Practitioner survey	Online pre and post intervention survey	65 settings	Survey analysis
		Interviews	Semi-structured interviews with practitioners, setting leaders, and trainer mentors	20 setting leaders (project group) 30 practitioners 7/8 trainer mentors.	Framework analysis

Responsiveness	1, 3, 4, 7, 8, 10, 13, 14	Practitioner survey	Online pre and post intervention survey	65 settings	Survey analysis
		Interviews	Semi-structured interviews with practitioners, setting leaders, and trainer mentors	20 setting leaders (project group) 30 practitioners (project group) 7/8 trainer mentors.	Framework analysis
Programme differentiation	14	Practitioner survey	Online pre and post intervention survey	130 settings	Survey analysis
		Interviews	Semi-structured interviews with practitioners, setting leaders, and trainer mentors	20 setting leaders (project group) 30 practitioners (project group) 7/8 trainer mentors. 10 setting leaders in control settings	Framework analysis
Monitoring of control	2, 14	Interviews	Semi-structured interviews with setting leaders (control)	10 setting leaders in control settings	Framework analysis
Perceived impact	4, 5, 7, 8, 10. 12, 13	Interviews	Semi-structured interviews with practitioners, setting leaders, and trainer mentors	20 setting leaders (project group) 30 practitioners (project group) 7/8 trainer mentors.	Framework analysis
Cost	5, 6, 7, 8, 11	Interviews	Semi-structured interviews with practitioners, setting leaders, and trainer mentors	20 setting leaders (project group). 30 practitioners (project group) 7/8 trainer mentors	Framework analysis

Context/moderators	4, 6, 7, 8, 9, 10, 12	Interviews	Semi-structured interviews with practitioners, setting leaders, and trainer mentors	20 setting leaders (project group) 30 practitioners (project group) 7/8 trainer mentors. 10 setting leaders (control group)	Framework analysis
Mediators	5, 6, 7, 8, 10, 12	Interviews	Semi-structured interviews with practitioners, setting leaders, and trainer mentors	20 setting leaders (project group) 30 practitioners (project group) 7/8 trainer mentors.	Framework analysis

Pre- and post-intervention surveys

Online pre- and post-intervention surveys will be carried out in all programme and control settings. Surveys in programme settings will collect data on programme fidelity, adaptation, responsiveness, programme differentiation, and contextual factors. Surveys in control settings will collect data on programme differentiation and contextual factors.

The surveys will be directed to the headteacher/manager of the nursery and frequent reminders will be used to ensure that the response rate is maximised. Ideally, all settings will complete the surveys, and the surveys will be designed to be sufficiently short to be completed in a reasonable time.

The surveys will include questions on:

- Fidelity: for example, in the post survey, what was the frequency of attendance of sessions by children and of professional development sessions by practitioners? What was the duration of programme delivery? Were the sessions consistently attended by the same children?
- Adaptation: for example, where any changes made to the way the programme was delivered?
- Responsiveness: for example, do practitioners perceive that children are engaging positively with the sessions?
- Programme differentiation: for example, does the setting offer other structured sessions to develop children's language skills? hat do they consist of? How often are they offered? Are they targeted or delivered in small groups?
- Contextual characteristics for example: type of settings, number of children eligible for extended entitlement (free 30 hours of childcare), number of EYPP-eligible children and EYPP funding, number of staff and qualifications.

Interviews

The interviews will provide in-depth data on experiences, views and perceived impacts. As the interviews are semi-structured, this flexible format gives the respondent the opportunity to introduce any themes that we have not originally anticipated. All interviews will last 30-45 minutes and will be conducted online or over the phone, depending on participant preferences.

In-depth interviews with practitioners will explore their perspectives around the training and resources provided, their experiences of programme delivery and their views around the perceived impact of *Talking Time*© on both their teaching practice and children's outcomes. The interviews will explore practitioners' views of each of the element of the intervention (the training sessions, mentoring sessions, individual professional reflection, resources and manuals for the structured activities), as well as the causal links between each activity and potential impact on teaching practices. They will also be asked about how they see the intervention contributing to the particular demographic of children that the setting serves. It will be important to explore how practitioners engage in adaptation and independent planning of activities throughout the programme, whether there is any variation in that and what factors impact practitioners' ability to do so. We propose a total of 30 interviews with practitioners (teachers, teaching assistants and other early years practitioners).

The possibility of interviews being requested will be included in the setting MOU, although of course, practitioners will have a right to refusal. We will send out a direct email to all intervention settings. We expect to have the names of the *Talking Time*© Leads, and so will be able to contact them directly. A clear accessible information sheet will explain the purposes of the research to practitioners, that we are interested in their perspectives and experiences and that there are not 'right' or 'wrong' answers to our questions, as well as the ethical guarantees of confidentiality, anonymity and right to withdraw. We will be clear that neither they nor the setting will be identifiable in any outputs. In the presentation of data, we will also remain alert to and guard against the possibility of colleagues recognising each other. The information sheet will also be accompanied by a request for volunteers, as settings will opt -in to the IPE. From amongst those that volunteer, we will balance a number of criteria. First we will ensure:

- A regional distribution of settings from which the practitioner interviews are drawn
- A mixture of school-based nursery settings and PVI settings

Once that is achieved, we will seek to ensure that a majority of the 30 settings serve high numbers of disadvantaged children and/or those with EAL (the focus for the intervention and research).

We will keep in contact with settings to remind them of our request. We are confident that 30 practitioners (teachers, teaching assistants and other early years practitioners) can be recruited in this way. However, if this does not lead to 30 volunteers, or if some of our criteria are under-represented, we will through direct (telephone) contact with particular settings try and ensure buy-in by explaining how important practitioner perspectives are to the evaluation and addressing any specific anxieties/queries.

We aim to conduct interviews that produce a representative profile of the study population. Hence, we would like to spread the interviews as evenly as possible across regions and type of setting. There are five geographic regions and two broad types of settings (maintained and PVI). For the practitioner interviews, we will identify six settings from each region, of which, as far as possible, three will be PVIs and three from maintained nurseries.

***In-depth interviews with all the trainer-*mentors** (expected to number 7/8 in total) from the delivery team will explore their experiences of delivering the training and mentoring sessions and their views on practitioner engagement and any factors influencing that, (specifically what are the benefits of the *Talking Time*© intervention, how outcomes can be enhanced, their views on any improvements, both in the intervention itself and in the training, which would benefit practitioners and children, what are the impediments and facilitators to implementation in different settings). Thus, these interviews will provide insights into the extent to which the programme was delivered as intended, and any potential barriers to delivery (including challenges/facilitators to training/mentoring programme and those to the programme of children's activities). Data collected in these interviews will respond to issues around fidelity, implementation, context and moderators, programme differentiation and perceived impact.

In-depth interviews with setting leaders (20 in the project group and 10 in the control group). The term 'setting leader' is used as a generic one that covers both headteachers / senior leaders in school settings, as well as managers in PVI settings. Setting leaders are included because they will have an overview on funding, staffing issues, the setting's pedagogical approach and the social and ethnic background of the families served by the setting, that individual practitioners directly engaged in the *Talking Time*© intervention may not have. The 20 setting leaders in the project group will be drawn, as far as possible, from the same settings where we have interviewed practitioners. As there is a smaller number of setting leaders involved, we will ask for volunteers initially from four of the six settings chosen in each of the five regions, focusing on those with higher numbers of disadvantaged children /children with EAL as these characteristics are particular points of interest to the evaluation. These interviews will explore the difference that the programme makes when compared to the settings' usual practice, as well as any impacts on longer-term leadership, planning, and adaptation. Interviews will also cover how the programme fits with the settings' wider professional development strategies, as well as any other training or development programmes that delivery staff might have engaged in. Setting leaders in the programme settings will be able to make in-depth comments on the impact of the intervention and the resulting changes made in pedagogy and practice in the settings, the experience of interacting with the trainer/mentors, and the way in which the intervention has (or not) embedded itself in the setting, and complements (or not) other professional development strategies. For the control settings, we will identify two settings from each region, using the criteria outlined above, and approach the setting lead to request an interview. The inclusion of the interviews with leaders in 10 control settings is to enhance understanding of business as usual in terms of settings' language development policy, pedagogy and practices.

Given the heterogeneity of the settings that is likely to be revealed through the IPE - in terms of organisation, pedagogy, and families served, as well as variations in individual practitioner engagement with *Talking Time*©, we do not expect saturation to be an issue. However,

researchers will stop interviewing if they realise at any point, after conducting at least 50% of interviews, that additional interviews are not bringing any additional information.

The interviews will be conducted between April 2025 – June 2025 by the evaluation team members, Sophie Kitson (Social Researcher), Jasmin Rostron (Social Researcher) and Ekaterina Aleynikova (Senior Social Researcher). All three are experienced qualitative researchers. The team will also analyse the data and draft the report. Carol Vincent will oversee all stages of the qualitative research and provide quality assurance. She is a highly experienced qualitative educational researcher.

Preparation study

Nineteen settings will be involved in delivery from January 2024. Five of those settings will be chosen for light-touch evaluation. The primary purpose of the preparation study is to pilot the interview schedules and the assessment process. Given that previous research on *Talking Time*© has involved maintained nursery settings, the preparation study will focus on PVI settings. The aim is to gather data on key research questions concerning fidelity and implementation. The RQs guiding the preparation study will be:

- Is the programme content accessible to practitioners in PVI settings? Do any elements need adapting?
- How do PVI practitioners and setting leaders perceive the programme?
- Are there barriers to delivery and engagement in PVI settings?
- Are there any unexpected challenges or benefits seen in PVI settings?

This will involve 2 interviews per setting, one with practitioners and one with setting leaders making a total of 10. These will be carried out online or over the phone as the respondents prefer in June 2024. The delivery team will identify PVI settings that are willing to take part in the prep study, and from within this grouping, we will then focus on settings with a high number of disadvantaged children and /or children with EAL. The data will be analysed through framework analysis (see below) and written up as an internal report for EEF by the same team as will work on the main study.

The evaluation team in collaboration with Qa will conduct a pilot study in five settings to assess the feasibility and acceptability of the proposed assessments. In particular, the pilot will investigate the time required to carry out the assessments; the difficulty and acceptability of each assessment as perceived by the assessor for each conducted test; whether the ordering of the assessments has an impact on the time required or on acceptability; whether feasibility varies in PVIs versus the maintained sector; whether assessment are more difficult to conduct with disadvantaged children.

The pilot will be conducted in 5 settings and will conduct a total of 50 assessments. During the pilot, the assessors will rate acceptability and difficulty on a scale from 1 to 5 for each child/assessment. In addition, assessors will make notes of the reasons for why some children do not take the assessment or stop taking the assessment at some point. Objective and measurable indicators of feasibility will include the proportion of children not completing the assessment and the average time taken to complete the assessments. Subjective

measures of feasibility will be based on the assessors' ratings of difficulty and acceptability mentioned above.

Follow-up study

'light-touch' follow-up study will take place in Spring 2026. The research questions for the follow-up study will be defined once the results of the main study are known. However, we can anticipate that in the follow-up study we will investigate questions relating to the long-term impacts of the intervention and its sustainability. In particularly, we will want to investigate to what extent *Talking Time*© has embedded in settings and continues to be followed by staff having produced permanent changes in their practice. In addition, we will collect information on any impact of the intervention beyond the endline study as perceived by staff. Settings will be approached for participation, the criterion being to ensure a varied range of longer-term reactions to *Talking Time*©.

The follow-up study will also explore issues of fidelity and adaptation. We will investigate to what extent the intervention is implemented in the way it was designed, what adaptations were made, and what are practitioners' attitudes towards the intervention to understand *why* the initiative was fully established, partially established (i.e. only some elements of *Talking Time*© remain in use) or abandoned.

At this stage, we suggest that the follow-up study will include:

- Returning to 10 settings that participated in *Talking Time*©;
- Conducting one interview per setting, of around 45 minutes in length (online or by telephone as participants prefer)
- Interviews will take place with the same interviewee who took part in the main study. If two interviewees (i.e. setting lead and practitioner) took part, we will leave it to the setting to determine who is most appropriate.

The data will be analysed through framework analysis (see below) and written up as an addendum to the main report by the same team as will work on the main study.

IPE Analysis

Qualitative data from the interviews, will be analysed to gather a rich picture of delivery, and the experiences and perspectives of practitioners and settings leads in relation to the *Talking Time*© initiative. All members of the evaluation team be involved in the analysis. The qualitative and quantitative analysis will inform each other during three workshops at key stages of the study as discussed above. The workshops will identify points of agreement and disagreement between the IE and the IPE, which will be reconciled with further qualitative analysis or exploratory quantitative analysis at later stages.

We will record all the interviews (with participant permission) and they will be transcribed verbatim by professional transcribers. The data will be analysed using a framework approach, a systematic method for categorising and organising qualitative data in order to be able to compare and contrast aspects of that data (Gale et al 2013), which is well designed for the

analysis of large data sets and compatible with *Nvivo* (Parkinson et al 2016). Framework analysis has five steps: familiarisation (with the data set, through close reading of the transcripts); identifying a thematic framework (the research team agreeing on emerging themes); indexing (identifying portions of the data that come under a particular theme); charting (bringing together data under each theme); and mapping/interpretation (developing a written analysis of the data). Framework analysis is 'better adapted to research that has specific questions, a limited time frame, a pre-designed sample (e.g. professional participants) and a priori issues (e.g. organizational and integration issues) that need to be dealt with. [It's] prime concern is to describe and interpret what is happening in a particular setting' (Srivasta & Thomson, 2009, p.73).

To build a thematic framework, we will use a combined inductive-deductive approach to coding an initial selection of transcripts, We will first use themes that draw upon prior knowledge from earlier research by Julie Dockrell and colleagues on *Talking Time*©, the IE dimensions, and the IPE research questions (i.e. a deductive approach). However, during data collection and analysis, we are likely to identify additional issues emerging from familiarization with the qualitative data, and will incorporate these into the framework, (i.e. a 'bottom-up' inductive approach). Our focus here will be on sifting and sorting the data rather than interpretation which will be a later focus (Parkinson et al. 2016).

The agreed codes will then be applied to the complete data set through an indexing of the transcripts, using a mixture of hand coding and *Nvivo*. In order to manage the volume of data, it will then be summarized and arranged into chart form by category, with specific examples of the data as illustration. The data will then be mapped (e.g. looking for patterns and discrepancies) and interpreted in a process of rigorous, transparent and data-driven meaning-making.

The quantitative data obtained from all survey settings will be used to compare baseline contextual characteristics between the programme and the control groups. We expect the two groups to be very similar as the result of randomisation, and these data will give us the opportunity to evaluate the success of the randomisation procedure by comparing standardised means of the two groups. Changes in contextual factors or changes in the presence of other programmes will be tabulated and compared descriptively across the two groups. The data on attendance of sessions by children and practitioners (number of sessions attended) will be presented descriptively in summary tables and employed in the compliance analysis. Attendance data will be used to assess to what extent the intervention was planned as intended in a quantitative way (fidelity). As such it will provide valuable information to understand the effectiveness of the intervention along the theory of change and will therefore support the interpretation of the results of the trial. Observed changes in staff attitudes and perceptions over the trial may also help the interpretation of the results of the trial.

The qualitative and quantitative analysis of the data will be conducted by different members of the team. The members of the evaluation team will meet at critical points of the evaluation (after the baseline, at midterm, and after the endline) to present the results to each other, and to discuss and reconcile any differences. Triangulation will occur through a discussion of the results in the workshops. Any discrepancies in the observed outcomes will be discussed and will lead to further areas of investigation (through further exploratory quantitative analysis of

the data and through the definition of new research questions for the qualitative research, particularly at the follow-up).

Cost evaluation design

We will follow the principles set out in the EEF cost evaluation guidance (EEF, 2023), and will estimate costs from the perspective of the setting and present these in the form of a per-child cost, calculated as if implemented over a three-year basis.

Costs will be separated into pre-requisites, start-up costs and recurring costs. A key cost of the programme is the staff time required. In line with EEF principles, practitioner time (for example, time spent in preparation, training and delivery) will be presented in the evaluation report in units of time. Throughout the cost evaluation, it will be important, as far as feasible, to identify additional costs required to implement the programme, compared to the costs that would be incurred as part of usual practice. In line with EEF guidance, costs will be based on those incurred in the trial, excluding any research costs and any subsidies provided by EEF or the SPHs for the purpose of the trial.

Information on costs will be obtained from the delivery team. Two online workshops will be held with members of the delivery team. During the first workshop the evaluation team will fill out the lists of resources and time use recommended by the EEF guidance on the collection of cost data (EEF, 2023). Additional information on resource use will be obtained from the surveys in programme settings conducted as part of the IPE, which will include questions on resource use. In the second workshop the teams will correct the figures in the resource table by addressing any pending question from the previous workshop.

We will use the ingredients approach in which costs of all resources required to implement the programme, regardless of who incurs the costs (but excluding research costs), are included. *Talking Time*© does not require additional staff or school hours and is conducted within the settings' premises, in such a way that the delivery does not imply additional costs. Ingredients of *Talking Time*© will include visits to the settings, professional development training, coaching sessions, and any follow-up activity with practitioners and the settings. Training occurring outside working hours or requiring cover will be accounted for at the cost of a cover for the time spent in the training, including travel or subsistence. The use of materials such as manual and other teaching supporting tools will be included at market prices.

To calculate cost-per-child, we will first estimate the total cost per setting if the programme were to be implemented for three consecutive years, and then work out the per year cost. To obtain a child-level cost estimate, the total cost will be divided by the number of children per setting who would be expected to benefit; we intend to use the actual number of children attending settings in the trial for this purpose. All assumptions used in the cost evaluation will be clearly stated in the report.

Ethics and registration

The proposal has been approved by the NIESR ethics committee after careful consideration and discussion of several issues regarding parental consent, namely accessibility of the documentation to some parents who have English as an additional language and the necessity of opt-in consent video/audio recordings. The delivery team suppled translated versions of

documents in a range of home languages for parents with EAL.. With regard to the issue of opt-in, it will be made clear on the parent information form that children will not be video/audio-recorded without parents' explicit consent. A consent form will be given to parents and only children whose parents have consented to their child participation in the study (including the possibility of audio recording of the assessments) will be included in the evaluation. Children will not be consulted directly because of their young age, but when assessed, the assessors will be alert to any signs of boredom and/or distress and if these are apparent, they will release the children to other activities

An information sheet for settings provides an overview of the project, and the first stage is for early years settings to express interest in the trial. Following this, to participate in the trial, settings will be asked to sign a memorandum of understanding, setting out in detail what participation in the project involves and the roles of all parties. Settings will then be asked to distribute information sheets to parents of eligible children, which will clearly explain the evaluation, and explain how parents can withdraw their child from the evaluation. The information pack will also include opt-in forms allowing parents to provide consent for their children participation in the study. For two elements of the project (audio recordings – which will only be conducted if the delivery team is able to secure additional funding for this element - and video recordings – to be used by the delivery team only), parents will be required to actively indicate consent for their child to participate in these activities, as noted above. Parents have a right to withdraw from the trial at any point with no detrimental consequences and data pertaining to that child will not be used. Practitioners will be given an information sheet and asked to give informed consent to participating in the project. Verbal information will be provided prior to any research encounter, and ethical consent will be confirmed. Verbal information will make clear the purpose of the research, how the data will be used, when the data will be deleted and how to withdraw data. Privacy notices including details of the data being collected will also be provided to settings, staff and parents.

The trial will be registered at the ISRCTN registry; the registration number will be added to the protocol once this becomes available. The record in the trial registry will be updated with outcomes at the end of the project.

Data protection

We recognise the utmost importance of data protection and are committed to complying with the Data Protection Act 2018 and the GDPR. The evaluation will involve the collection of personal data in relation to participating children, as well as personal data relating to participating staff in early years settings.

A DPIA has been conducted for the evaluation by the evaluation team, which has been reviewed by IESR's D O and will be kept under review through the duration of the project. A separate DPIA is also being conducted by the delivery team.

The evaluation team's legal basis for processing personal data is legitimate interests. NIESR considers that it has a legitimate interest in processing the personal data in order to conduct the *Talking Time*© evaluation (with undertaking research and evaluation activities one of IESR's core purposes), the processing identified is necessary in order to meet this purpose,

and there are minimal impacts of the processing on the individuals involved (and which we have taken steps to ensure).

The special category data to be processed for this evaluation relates to gender and to languages spoken in the home (which could potentially be indicative of ethnicity). IESR's legal basis for processing special categories of personal data is rticle 9j) 'archiving, research and statistics.' The evaluation of *Talking Time*© is particularly concerned with interventions that can support language skill development of children with EAL. Processing personal information to home language is necessary for the conduct of the evaluation.

A data sharing agreement (DSA) is being agreed between all parties involved, which states which data will be shared by who, how and why.

A privacy notice, drafted in collaboration with the delivery team and EEF, explains how information collected from participants will be used and stored, and communicates to participants their right to withdraw from data processing. All participants interviewed for the research will be asked for their consent to indicate that they understand the aims of the research and agree to the interview being recorded and transcribed, and given assurance of anonymity. Settings will sign an MOU at the start of the project clearly laying out the requirements of participation and how data will be used, shared and stored. Settings completed a data sharing agreement which covered the data processing of video and audio recordings.

The privacy notice also explains that the personal data described may be shared with third parties. These third parties may include the *Talking Time*© mentors, trusted academic partners and subcontractors who may process information on behalf of the evaluation and delivery teams to carry out work such as transcribing interviews and aspects of programme delivery. In all cases, the evaluation and delivery teams will ensure that these third parties enter into appropriate data processing agreements and that they keep the personal data secure and confidential.

The evaluation team, and the delivery team will be data controllers of the personal data collected during the evaluation, and our assessment partner (Qa Research) will be data processors. At the end of the project (once the evaluation report is finalised), the evaluation team will upload the child outcome data to the EEF archive; this will take place through the ONS Secure Research Service. Once this data is in the archive, the EEF will become the data controller and the Fischer Family Trust, (FFT) will be the data processor for this data. For the purposes of research and archiving, the responses will be linked with information about the children from the National Pupil Database (NPD) and shared with the Department for Education, the EEF, the EEF's archive manager, the Office for ational Statistics and potentially other research teams. NIESR will delete personal data held in relation to the project six months following publication of the final evaluation report.

Personnel

Evaluation team (NIESR)

Edoardo Masset (Associate Research Director, NIESR) is the principal investigator. He will be responsible for the overall evaluation study. He will lead the evaluation team, design the study, lead and carry out the analysis of the data, and write the final report. He will also support the coordination between the evaluation team, the delivery team, and the data collection team. Edoardo is an expert of impact evaluation methods. He has been the technical lead of several experimental and quasi-experimental evaluation and has great experience in managing evaluations of complex interventions, Prior to joining NIESR he was associate professor at the London School of Hygiene and Tropical medicine, where he led a an FCDO-funded programme promoting and supporting the development and use of new evaluation methods in challenging environments.

Ekaterina Aleynikova (Senior Social Researcher, NIESR) will contribute towards all aspects of the qualitative data collection and analysis (interviews, analysis and writing up). She is a qualitative researcher with extensive experience of doing research with school workforces. She is currently leading IESR's research into the issues impacting recruitment and retention of senior school leaders in Wales, which involves in-depth interviews and focus groups with senior and middle school leaders. Prior to joining NIESR, Ekaterina worked at Learning and Work Institute where she successfully managed and delivered multiple high quality research projects.

Sophie Kitson (Assistant Social Researcher, NIESR), with other NIESR colleagues, will be responsible for conducting interviews with practitioners, setting leads and trainer/mentors. She will also contribute to analysis and writing the reports. Sophie is predominantly a qualitative researcher, with experience in interviews, focus groups, content analysis and reporting. She has experience in recruiting and conducting in-depth interviews with the school workforce, including school leaders, middle leaders, teachers and support staff. Her recent work includes analysis and reporting for a mixed-methods evaluation of the What Works Centre for Children's Social Care intervention, focusing on Designated Safeguarding Leads in schools. She is also working on IESR's research into the recruitment and retention of senior school leaders in Wales.

Jasmin Rostron (Associate Social Researcher, NIESR), with other NIESR colleagues, will be responsible for conducting interviews with practitioners, setting leads and trainer/mentors. She will also contribute to analysis and writing the reports. Jasmin is a qualitative researcher, experienced in interviews, focus groups and evaluations. She has led on two implementation and process evaluations for hat orks for Children's Social Care (CSC) trials in schools. This involved interviewing designated safeguarding leads, middle and senior leaders. Additionally, she has worked with Refugee Education UK on an ONS-funded qualitative project, on the experiences of displaced young people, accessing education and healthcare.

Carol Vincent (Associate Research Director (Education and Skills), NIESR**)** will provide quality assurance on the project. Carol is a qualitative researcher with more than twenty years' experience of managing qualitative funded research projects in the university sector, and

conducting interviews with the education workforce, including early years practitioners. She has an extensive track record in writing for publication and for a range of different audiences.

Claudine Bowyer-Crane (NIESR Fellow and Professor at the University of Sheffield) Claudine is a psychologist with an interest in children's reading and language development, particularly children learning English as an Additional Language. She has extensive expertise in designing and evaluating interventions to support children's language development. Claudine will advise the team on the identification, use, and analysis of language skills indicators.

Richard Dorsett (NIESR fellow) Richard is a Professor of Economic Evaluation at the University of Westminster. His research interests include programme evaluation, labour market transitions, education, and well-being. He specialises in empirical research and the application of econometric techniques to estimate causal impacts.

Delivery team (University of Oxford and UCL)

Sandra Mathers (Associate Professor, University of Oxford) is the principal investigator for the Implementation Team. She will be responsible for leading the intervention delivery and team, providing intellectual and organisational leadership. Sandra is an expert in professional development and early intervention within disadvantaged populations. Her research explores how we can best promote high-quality interactions between adults and children, including developing and evaluating early language and professional development programmes (U*RLEY, Talking Time©, Texts for Talk*); assessing the impact of early childhood provision and adult-child interactions within large-scale longitudinal studies (M*illennium Cohort Study, Children of the 2020s, Evaluations of the Graduate Leader Fund and Early Education Pilot for Two-Year Olds*); and studying and quality and inequality in early education provision. She was co-investigator on the recent randomised controlled trial of *Talking Time*© funded by the Nuffield Foundation, leading on the development of the professional development component/manual and programme delivery.

Julie Dockrell (UCL) is Professor of Psychology and Special Needs at the UCL, Institute of Education and a co-investigator for the Implementation Team. She will provide intellectual leadership and expertise on child oracy. Julie developed the original *Talking Time*© programme and led its first evaluation in a disadvantaged area. She was principal investigator on the recent randomised controlled trial of *Talking Time*©, funded by the Nuffield Foundation. Julie is a leader in the assessment of children's oracy and has led wide-ranging national and international research.

Liz Hewitt (Research Officer, University of Oxford) is part of the implementation team. She will support the organisation and management of *Talking Time*© delivery. In addition, Liz will lead on recruitment of settings, liaising with Stronger Practice Hubs. Liz's academic research background lies in qualitative study of teachers' use of pupil discussion. Liz joined the *Talking Time*© team during the recent randomised controlled trial of *Talking Time*©, funded by the Nuffield Foundation. She supported the implementation and process evaluation work. Prior to joining University of Oxford, Liz worked in Initial Teacher Education and as a primary teacher. She has also worked as a Researcher with Ofsted.

Risks

Table 6 Risks

Issue/risk	Action to address issue/reduce risk	Likelihood	Impact
Reluctance of settings to participate	The delivery team will primarily be responsible for recruitment although we will work closely with them, and ensure sufficient time to recruit settings will be key, as well as providing clear and accessible recruitment materials.	Medium	High
Settings do not provide the data required	Responsibilities will be clearly laid out in an MoU. We will work to reduce burdens on settings with clear guidelines on requirements, with long lead in times to help planning.	Medium	High
Attrition, especially non-completion of post-tests	Requirements of the study, including assessments, will be made clear to settings in recruitment. Ensuring flexibility in the scheduling of assessments and minimising burden for settings where possible. Incentives are also to be provided which should also help retain participation, especially among the control group.	Medium	High
Large dropout rates in settings with poor capacity.	The data collection team will conduct mopping-up visits in order to administer the largest possible number of assessments at both baseline and follow-up. Dates of assessments will be shared with parents and practitioners well in advance, particularly in more difficult settings.	Medium	High
Data collection process does not adhere to agreed protocol	The evaluation team will meet periodically with the data collection team to define a feasible data collection plan. A member of the evaluation team will make spot visits to settings where data collection take place to check adherence to the protocol. This will occur in the first week of data collection in order to be able to make any remedial correction to the protocol.	Low-medium	High
Reluctance of setting staff to participate in interviews	Requirements will be made clear to settings during recruitment. We will ensure flexibility in scheduling interviews to accommodate setting staff timetables, and emphasise that we are wishing to explore staff experiences and perspectives (therefore no 'right' or 'wrong' answers).	Low-Medium	Medium
Reluctance of control group setting leads to participate in interviews	We are only seeking 10 control group setting lead interviews. We expect that the control group setting leads are interested in the TT	Low-Medium	Medium

	intervention (they may choose to spend the £1000 payment on buying it) and therefore some setting leads would be willing to participate in a 'business as usual' interview, contributing to the evaluation.		

Timeline

Table 7: Timeline

Dates	Activity	Detail / Scheduled dates	Staff responsible/ leading
July-Aug 2023	Set up meetings and IDEA workshop	1 August: IDEA workshop 4 August: 1st Set-up meeting 22 August: 2nd Set-up meeting	Delivery team and evaluation team
Sept – Nov 2023	Ethical approval; develop recruitment materials (including for evaluation preparation study)	22 October: Application for ethical approval submitted to NIESR ethics committee 11 December: Ethics application form approved by NIESR ethics committee	Evaluation team and delivery team
Dec 2023-July 2024	Recruitment of early years settings (recruitment for both pre-trial and main trial starts in January 2024)	Ongoing recruitment led by delivery team and Stronger Practice Hubs (SPH)	Delivery team with support from evaluation team
Jan 2024 – September 2024	Subsidised delivery period prior to main trial	February 2024: Rolling start date for delivery of programme in settings, subject to when settings sign on for preparation study.	Delivery team
April 2024 – July 2024	Evaluation activities as part of evaluation preparation study (interviews and test-run of assessments)	April 2024: Evaluation team contact settings to organise interview dates with EY staff Given the rolling date of onboarding, the delivery team will assist in selecting EY settings for assessments to allow for as high dose to the programme as possible.	Evaluation team

		June-July 2024: Qa research conduct 50 pilot assessments in 5 EY settings agreed with Oxford University. Interim findings are shared during a planned session with the delivery team in order to improve programme delivery.	
Mar 2024 – July 2024	Reporting on findings of evaluation preparation study (presentation plus ongoing feedback meetings)	March: Study protocol. Submit protocol, revise based on feedback and publish study protocol. 8th July: Data from pilot assessments and interviews compiled, analysed, and draft report of findings shared with EEF, UCL, and Oxford University.	Evaluation team
August – October 2024	Pre-intervention data collection and baseline assessments	By 2nd September: Data collection template shared with EY settings and settings supported to distribute information for parents and consent forms (parents have 2 weeks to withdraw) 23 September – 25 October: Baseline assessments	Evaluation team
End October 2024 (half-term)	Randomisation	28 October – 1 November: Randomisation	Evaluation team
November – December 2024	Mop-up assessments	4 November – 20 December: Qa research revisits EY settings to conduct mop-up assessments	Evaluation team
January 2025	Statistical Analysis Plan (SAP)	Submit draft Statistical Analysis Plan (SAP)	Evaluation Team
January 2025 – March 2025	IPE preparation and planning	Interviews guides and semi-structured questionnaires finalised, settings contacted for the interviews and interviews scheduled	Evaluation Team

December 2024 – June 2025	Intervention delivery (20 weeks) - includes welcome visits/briefings	November/December 2024: welcome visits to settings 6 January: Group 1 EY staff participate in professional development training and commence Talking Time delivery 13 January: -; Group 2 EY staff participate in professional development training and commence Talking Time delivery 20 January: ; Group 3 EY staff participate in professional development training and commence Talking Time delivery 27 January:	Delivery team
March 2025	Statistical Analysis Plan (SAP) IPE design	SAP published on EEF website Topic schedules for IPE completed. Prepare for IPE data collection.	Evaluation team
April 2025 – June 2025	IPE interviews	IPE data collection Obtain attendance data	Evaluation team
May – July 2025	Endline assessments	9 June – 27 June: Qa research conducts endline assessments with Group 1 16 June – 4 July: Qa research conducts endline assessments with Group 2 23 June – 11 July: Qa research conducts endline assessments with Group 3	Evaluation team
July - November2025	Analysis and reporting	July- August: QA and analysis of IPE and IE data. August- November: reporting	Evaluation team
September 2025	Cost analysis	2 online workshops to compile all costs incurred by the project	Evaluation team and Delivery team

November 2025	First draft of evaluation report	Submit first draft of evaluation report	Evaluation team
January - March 2026	Follow up to IPE: Returning to 10 settings that participated in *Talking Time*©; conducting one interview per setting, of around 45 minutes in length Evaluation report published in EEF website		Evaluation team
April 2026	Draft of report on follow up	Submit draft of addendum report	Evaluation team
May 2026	Final report on main trial, including report on follow up as an appendix	Final report published	Evaluation team
July 2026	Archiving data	Submit data to EEF archive	Evaluation team

References

August, D. and Shanahan, T. (2006) *Developing literacy in second language learners: Report of the national literacy panel on language minority children and youth*. Edited by D. August and T. Shanahan. Mahwah, NJ: Erlbaum.

Borenstein, M. *et al.* (2009) *Introduction to Meta-Analysis*, *Introduction to Meta-Analysis*. Chichester, UK: John Wiley & Sons, Ltd. Available at: https://doi.org/10.1002/9780470743386.

Burgoyne, K. *et al.* (2 18) 'Evaluation of a parent-delivered early language enrichment programme: evidence from a randomised controlled trial', *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 59(5). Available at: https://doi.org/10.1111/jcpp.12819.

Dobinson, K.L., Mathers, S. and Dockrell, J.E. (2022) *The quality of universal language support in mantained nursery classrooms in two socially disadvantaged areas of England*, *Unpublished manuscript*.

Dockrell, J.E. *et al.* (2 17) 'Children with speech language and communication needs in england: challenges for ractice', *Frontiers in Education*, 2. Available at: https://doi.org/10.3389/FEDUC.2017.00035.

Dockrell, J.E. *et al.* (2023) *Empowering staff to enhance oral language in the early years: Cluster randomised trial: Final report*. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/eef-guidance-reports/literacy-early-years/Preparing_Literacy_Guidance_2018.pdf?v=1706592129 (Accessed: 31 January 2024).

Dockrell, J.E., Stuart, M. and King, D. (2 1) 'Supporting early oral language skills for English language learners in inner city preschool provision', *British Journal of Educational Psychology*, 80(4). Available at: https://doi.org/10.1348/000709910X493080.

Durham, R.E. *et al.* (2 7) 'Kindergarten oral language skill: key variable in the intergenerational transmission of socioeconomic status', *Research in Social Stratification and Mobility*, 25(4). Available at: https://doi.org/10.1016/j.rssm.2007.03.001.

Education Endowment Foundation (2018) *Preparing for Literacy: Improving Communication, Language and Literacy in the Early Years*. London.

EEF (2022a) *Implementation and process evaluation guidance for EEF evaluations*. London: EEF.

EEF (2022b) *Statistical Analysis Guidance for EEF Evaluations*. London: EEF.

EEF (2023) *Cost evaluation guidance for EEF evaluations*. London: EEF.

Efron, B. and Hastie, T. (2016) *Computer Age Statistical Inference*, *Computer Age Statistical Inference*. Available at: https://doi.org/10.1017/cbo9781316576533.

Gardner, R. (2023) *The development and evaluation of a preschool language enrichment programme: Can a practitioner-delivered, multi-tiered, short-term intervention improve early language outcomes?* University of Oxford.

Golinkoff, R.M. *et al.* (2 19) ‘Language Matters: Denying the Existence of the -Million-Word Gap Has Serious Consequences’, *Child Development*. Available at: https://doi.org/10.1111/cdev.13128.

González, S. and Bonal, X. (2 21) ‘CO ID-19 school closures and cumulative disadvantage: Assessing the learning gap in formal, informal and non-formal education’, *European Journal of Education*, 56(4). Available at: https://doi.org/10.1111/ejed.12476.

Hassinger-Das, B. *et al.* (2 17) ‘ matter of principle: pplying language science to the classroom and beyond.’, *Translational Issues in Psychological Science*, 3(1). Available at: https://doi.org/10.1037/tps0000085.

Hoff, E. (2 14) ‘Causes and consequences of SES-related differences in parent-to-child speech’, in *Socioeconomic Status, Parenting, and Child Development*. Available at: https://doi.org/10.4324/9781410607027-15.

Huttenlocher, J. *et al.* (2 1) ‘Sources of variability in children’s language growth’, *Cognitive Psychology*, 61(4). Available at: https://doi.org/10.1016/j.cogpsych.2010.08.002.

ICAN and RCSLT (2018) *Bercow: 10 years on. An independent review of provision for children and young people with speech, language and communication needs in England.* London.

Kieffer, M.J. (2 8) ‘Catching Up or Falling Behind? Initial English roficiency, Concentrated overty, and the Reading Growth of Language Minority Learners in the United States’, *Journal of Educational Psychology*, 100(4). Available at: https://doi.org/10.1037/0022-0663.100.4.851.

Law, J. *et al.* (2 9) ‘Modeling developmental language difficulties from school entry into adulthood: Literacy, mental health, and employment outcomes’, *Journal of Speech, Language, and Hearing Research*, 52(6). Available at: https://doi.org/10.1044/1092-4388(2009/08-0142).

Law, J. *et al.* (2 18) ‘Early Home ctivities and Oral Language Skills in Middle Childhood: Quantile nalysis’, *Child Development*, 89(1). Available at: https://doi.org/10.1111/cdev.12727.

Law, J., Charlton, J. and smussen, K. (2 17) ‘Language s a Child ellbeing Indicator’, *Early intervention foundation* [Preprint], (SEPTEMBER).

van der Lely, H.K.J., ayne, E. and McClelland, . (2 11) ‘ n investigation to validate the grammar and phonology screening (GAPS) test to identify children with specific language impairment’, *PLoS ONE*, 6(7). Available at: https://doi.org/10.1371/journal.pone.0022432.

Levine, D. *et al.* (2 2) ‘Evaluating socioeconomic gaps in preschoolers’ vocabulary, syntax and language process skills with the Quick Interactive Language Screener (QUILS)’, *Early Childhood Research Quarterly*, 50. Available at: https://doi.org/10.1016/j.ecresq.2018.11.006.

Massonnié, J. *et al.* (2 22) ‘Oral language at school entry: dimensionality of speaking and listening skills’, *Oxford Review of Education* [Preprint]. Available at: https://doi.org/10.1080/03054985.2021.2013189.

Morra ellegrino, M.L. and Scopesi, . (199) ‘Structure and function of baby talk in a day-care centre’, *Journal of Child Language*, 17(1). Available at: https://doi.org/10.1017/S030500090001312X.

Nelson, K.E. *et al.* (2 11) ‘Language delays of impoverished preschool children in relation to early academic and emotion recognition skills’, *First Language*, 31(2). Available at: https://doi.org/10.1177/0142723710391887.

Ofsted (2020) *Education recovery in early years providers: summer 2022*. Ofsted. Available at: https://www.gov.uk/government/publications/education-recovery-in-early-years-providers-summer-2022/education-recovery-in-early-years-providers-summer-2022 (Accessed: 31 January 2024).

Pace, A. *et al.* (2 17) ‘Identifying pathways between socioeconomic status and language development’, *Annual Review of Linguistics*. Available at: https://doi.org/10.1146/annurev-linguistics-011516-034226.

Romeo, R.R. *et al.* (2 18) ‘Beyond the -Million- ord Gap: Children’s Conversational Exposure Is Associated With Language-Related Brain Function’, *Psychological Science*, 29(5). Available at: https://doi.org/10.1177/0956797617742725.

Roulstone, S. *et al.* (2 1) ‘Investigating the role of language in children’s early educational outcomes’, *Department for EducationEducation* [Preprint].

Rowe, M.L. (2 22) ‘Environmental influences on early language and literacy development: Social policy and educational implications’, in *Advances in Child Development and Behavior*. Available at: https://doi.org/10.1016/bs.acdb.2022.04.001.

Rowe, M.L. and Snow, C.E. (2 2) ‘ nalyzing input quality along three dimensions: Interactive, linguistic, and conceptual’, *Journal of Child Language*, 47(1). Available at: https://doi.org/10.1017/S0305000919000655.

Shanahan, T. (2 6) ‘Relations among Oral Language, Reading and riting Development’, in *Handbook of Writing Research*.

Singh, A. *et al.* (2023) *Improving power calculations in educational trials*. Available at: https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluation/eef-evaluation-reports-and-research-papers/methodological-research-and-innovations/improving-power-calculations-in-educational-trials (Accessed: 31 January 2024).

asik, B. (2 8) ‘ hen fewer is more: Small groups in early childhood classrooms’, *Early Childhood Education Journal*, 35(6). Available at: https://doi.org/10.1007/s10643-008-0245-4.

eiler, B.K. and Decker, .L. (2 22) ‘The Impact of SES on Language Domain in Kindergartners’ Quick Interactive Language Screener (QUILS) erformance’, *Communication Disorders Quarterly*, 43(2). Available at: https://doi.org/10.1177/15257401211017475.

Wright, H. *et al.* (2020) *Using Research Tools to Improve Language in the Early Years: Evaluation report and executive summary*. London.

Appendix 1

The GAPS sentence repetition assessment

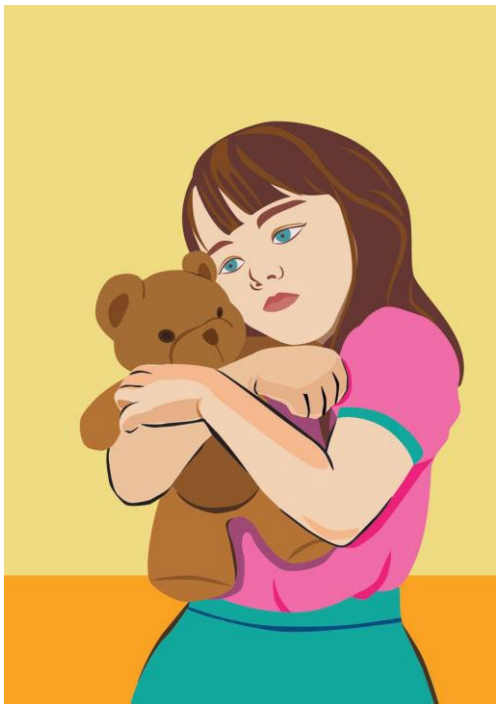
1. Sentence Repetition Score Sheet

Correct score = all words repeated correctly
OR no mistakes on **bold** words / parts of words

Sentence *Circle any words or parts of words not repeated correctly*	**Response** *Write here what the child said instead of the circled words*	✓	×
Practice Sentences			
a) This is the cat. *(Repeat if incorrect)*		FILLER	
b) The cat is grey. *(Repeat if incorrect)*		FILLER	
Test Sentences			
1. **The** cat **with the** bell **is** happy.			
2. The cat want**ed** some milk.			
3. The milk **is pulled by** the dog.			
4. *You say to Bik*: **What will** the cat drink?			
5. Not milk!		FILLER	
6. *You say to Bik*: **Which** dog **did** the cat push?			
7. The blue dog.		FILLER	
8. The dog gives **the cat the milk**.			
9. This cat like**s** milk.			
10. The cat is washing **herself**.			
11. The dog **is licked by** the cat.			
12. The cat is washing **him**.			
13. *You say to Bik*: **Who did** the cat wash?			
14. The red dog.		FILLER	
	TOTAL:		

A sample of the RAPT test (cards 1 and 2 of 10)

Card 1 What is the girl doing?



Card 2 What is the mother going to do?

Appendix 2 Stata randomisation code

```
/*fabricate a dataset of 130 settings observations with 8 strata of equal size*/
set seed 10000
drop _all
set obs 130
generate rand = uniform()
egen strata = cut(rand), group(8)
gen id=_n
drop rand

/*randomise within strata*/
generate rand=uniform()          /*generate a random number*/
sort strata rand                 /*sort observations by the random number within strata*/
egen treated = seq(), f(1) t(0)  /*assign every other observation to the intervention
group*/

/*test the two groups are similar*/
ttest rand, by(treat)
```