Education Endowment Foundation

Efficacy trial of Talking Time©: An oral language intervention for early years

Evaluation report

Originally published July 2026
Updated October 2026

Edoardo Masset, Anisa Butt, and Sherelle Davids

National Institute of Economic and Social Research

The Education Endowment (EEF) Foundation is an independent charity dedicated to breaking the link between family income and education achievement. We support schools, colleges, and early years settings to improve teaching and learning for 2 to19-year-olds through better use of evidence.

We do this by:

- **Summarising evidence**. Reviewing the best available evidence on teaching and learning and presenting in an accessible way.
- **Finding new evidence.** Funding independent evaluations of programmes and approaches that aim to raise the attainment of children and young people from socio-economically disadvantaged backgrounds.
- **Putting evidence to use**. Supporting education practitioners, as well as policymakers and other organisations, to use evidence in ways that improve teaching and learning.

We were set-up in 2011 by The Sutton Trust partnership with Impetus with a founding £125m grant from the Department for Education. In 2022, we were reendowed with an additional £137m from government, allowing us to continue our work until at least 2032.

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

Education Endowment Foundation
5th Floor, Millbank Tower,
21–24 Millbank,
London,
SW1P 4QP

0207 802 1653

info@eefoundation.org.uk

www.educationendowmentfoundation.org.uk

Table of contents

Executive summary **4**

Introduction **6**

Methods **18**

Impact evaluation results **38**

Implementation and process evaluation results **59**

Conclusion **90**

References **95**

Appendix A: EEF cost rating **99**

Appendix B: Security classification of trial findings **100**

Further Appendices **102**

About the evaluator

The project was independently evaluated by a team from the National Institute of Economic and Social Research including Edoardo Masset, Anisa Butt, and Sherelle Davids. Edoardo Masset was the project lead.

Contact details:

Edoardo Masset
National Institute of Economic and Social Research
2 Dean Trench Street,
London,
SW1P 3HE

Email: e.masset@niesr.ac.uk

Acknowledgements

We would like to thank the staff and all participating nurseries for their collaboration, as well as the children and their families for taking part in the study. The project was implemented thanks to the efforts of the mentor team led by Wendy Lee and Clare Williams, including Jenny Berry, Carla Cornelius, Sarah Mezey, Liz Stratton, Elinor Warner, and Helen Wood. The following National Institute of Economic and Social Research staff contributed to the evaluation and to the final report in various ways and at different stages: Carol Vincent; Sophie Kitson; Joanna Fang; and Marley Williams. The project team supported the study with comments in an advisory role all along the project and included Claudine Bowyer-Crane (University of Sheffield), Julie Dockrell (University College London), Ignacia Valenzuela (the Education Endowment Foundation [EEF]), and Aoife Duff (EEF). We would also like to thank the delivery team for their continuous support throughout the project, in particular Elizabeth Hewitt, Sandra Mathers, and Keeley Dobson. Finally, we would like to thank the staff at Qa Research, especially Katie Morris and Rachel Brown, for managing data collection, and about 60 education assessors who conducted the assessments in nurseries.

This report was published in July 2026 and updated in October 2026 to reflect minor wording changes.

Executive summary

The project

Talking Time© is an oral language development programme designed to improve the English language skills of children aged three to five years, with a particular focus on supporting children eligible for Early Years Pupil Premium (EYPP) and children with English as an Additional Language (EAL). The programme was delivered universally to all children aged three to four years in nursery settings. Children took part in two 15-minute small-group sessions per week over a 20-week period, focusing on structured oral language activities including shared storytelling, playful language games, and narrative discussions of familiar routines. Sessions were delivered by specifically trained early years setting staff (usually two leads per setting in the current study), supported by a structural professional development programme that included training, mentoring, and ongoing reflective practice. Talking Time© was originally developed by the Institute of Education at University College London (UCL) and subsequently expanded by the University of Oxford, which integrated the child-focused activities with a comprehensive professional development programme.

The impact of Talking Time© was evaluated through a cluster randomised controlled trial (RCT) involving approximately 1,700 children aged three to four years across 123 settings, including maintained nursery schools and private, voluntary and independent (PVI) providers in England. This was an efficacy trial conducted under supported delivery conditions. Children's language outcomes were measured at baseline and at follow-up using–Renfrew tests (expressive vocabulary and action picture), and the Grammar and Phonology Screening (GAPS) test. Alongside the impact evaluation, an implementation and process evaluation (IPE) was conducted, drawing on practitioner surveys and semi-structured interviews with practitioners, setting leads, and trainer-mentors to examine fidelity, implementation, adaptation, and perceived impact. The trial ran from autumn 2024, with baseline assessments conducted between September and November 2024, intervention delivery from November 2024 to June 2025, and follow-up assessments completed by July 2025.

As part of the Department for Education's Early Years Recovery Programme, the Education Endowment Foundation (EEF) worked with Stronger Practice Hubs (SPHs) across England to fund early years settings' access to evidence-informed programmes and study the programme's influence on practice and children's outcomes. The EEF worked with Liverpool City Region and Beyond, A Brighter Start, REACHout, and Thrive Together Early Years SPHs to fund settings' access to Talking Time© and evaluate the programme through this efficacy trial.

Table 1: Key conclusions

Key conclusions
1. Children in Talking Time© settings made one month's additional progress in language development, on average, compared with children in other settings. This result has a high security rating.
2. Among children eligible for EYPP, those in Talking Time© settings made approximately two additional months' progress in language development compared to those in other settings. Similarly, children with EAL made around two extra months' progress in English language development relative to those in other settings. However, these results may have lower security than the overall findings because of the smaller number of children analysed.
3. Among the different components of language development assessed, Talking Time© appeared to have the strongest impact on expressive vocabulary. There was less evidence of an impact on grammar overall, although subgroup findings suggest the programme may have been particularly beneficial for grammar skills among children with EAL and children eligible for EYPP.
4. Implementation of Talking Time© largely occurred as planned. The programme was well attended by children (80% of children attended at least 32 sessions—80% of the programme), and variation in attendance did not affect average programme impact.
5. Talking Time© is designed to be embedded in standard nursery practice and sustained independently. Practitioners reported benefits from the professional development, perceived improvements in children's oral language, and confidence in continuing delivery. However, challenges around staff capacity, particularly turnover, and limited flexibility in timetabling and session content suggest further adaptation and training may be needed to sustain delivery in years two and three.

EEF security rating

These findings have a high security rating. This was an efficacy trial, which tested whether the intervention worked under developer-led conditions in a number of settings. The trial was a well-designed two-arm RCT. The trial was well-powered. Children in Talking Time© settings were similar to those in the comparison settings in terms of prior attainment. However,

the following factors reduced the security of the trial. Around 15.6% of the children who started the trial were not included in the final analysis mainly because of child absence on testing days and relocation to different settings.

Additional findings

Children in Talking Time© settings made one month's additional progress, on average than those in the control group equivalent. This is our best estimate of impact, which has a high security rating. As with any study, there is always some uncertainty around the result: the possible impact of this programme also includes no additional progress and up to two months' progress.

The programme developers expected children eligible for EYPP and children with EAL to benefit most, given their lower baseline English language proficiency (Dockrell *et al.*, 2010). The findings are broadly consistent with this: estimated effects were larger for both groups than for the full sample. However, these subgroup results should be interpreted cautiously. Confidence intervals were wide and sample sizes small, particularly for EYPP children. Interaction analyses also did not provide statistical evidence that effects differed from those for other children. Qualitative evidence from the IPE suggests the intervention may have been particularly beneficial for children with language difficulties, but this finding is indicative.

Exploratory analyses suggest larger effects in maintained settings than in PVIs, although this is based on a small sample and should be interpreted with caution. This may reflect implementation differences identified in the IPE, including lower dosage in PVIs due to more variable attendance, and usual practice in these settings being closer to the programme's approach.

Further aims of the intervention were to embed Talking Time© within everyday nursery practice and strengthen practitioners' confidence and skills. Practitioners expressed strong enthusiasm for the programme and a willingness to continue delivering it in future years. The majority of respondents reported high levels of satisfaction with the professional development, and improvements in understanding, and confidence and skills in supporting early language development. Changes in practices extended beyond formal sessions and some elements became embedded with everyday practice. Overall, the findings suggest that the programme supported sustained changes in practice. However, staff also highlighted some challenges including limited staff time for planning delivering Talking Time© sessions and high staff turnover.

Follow-up interviews conducted the year after the trial provided encouraging evidence of sustainability, with most settings (seven out of nine) continuing to implement the programme and reporting longer-term changes in practices. Although this finding should be considered as indicative rather than conclusive given the small sample size. In addition, two of the settings in the follow-up sample had discontinued implementation for reasons that point to the importance of establishing mechanisms for training new staff to support the long-term sustainability of the intervention.

Overall, these findings are consistent with previous studies of Talking Time©. A small-scale study in three nurseries in a disadvantaged London borough found improvements in vocabulary and some other aspects of oral language, but more limited effects on narrative skills (Dockrell *et al.*, 2010). A later Nuffield Foundation randomised trial involving 36 settings found positive effects on expressive vocabulary, but not on other aspects of language development (Dockrell *et al.*, 2023).

Cost

The average cost of Talking Time© for one setting was around £3,505.52, or £64.25 per child per year when averaged over three years. This estimate includes 20 hours of staff time per setting to attend Talking Time© training. Practitioners reported that delivering the programme required additional staff time, however, this was not costed in the analysis. The cost of the programme was driven primarily by training and programme management activities. These costs were largely fixed across settings and, to some extent, independent of setting size.

Impact

Table 2: Summary of impact on primary outcome

Outcome / group	Effect size (95% confidence interval)	Estimated months' progress	EEF security rating	No. of children	P-value	EEF cost rating
Language development	0.08 (0.00,0.16)	1 month		1,439	0.047	£ £ £ £ £
Language development, EYPP-eligible children	0.10 (-0.04,0.25)	2 months	N/A	256	0.168	N/A

Introduction

Background

Children from economically disadvantaged backgrounds experience disproportionate delays in their language development compared to their more advantaged peers. Evidence consistently shows that children from poor socio-economic backgrounds are more than twice as likely to have speech, language, and communication needs (I CAN and RCSLT, 2018; Law *et al.,* 2009; Romeo *et al.,* 2018). These gaps emerge early and tend to widen over time with long-term implications for children's academic progress, mental health, and future employment (Law *et al.,* 2009; Roulstone *et al.,* 2010). The COVID-19 pandemic has intensified pre-existing inequalities: disruptions to early education have disproportionately affected children in deprived areas (González and Bonal, 2021). A recent report by the Office for Standards in Education, Children's Services and Skills (Ofsted) (2022) 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. This evidence points to the need for early years interventions targeted at settings serving high proportions of children at risk of language delay.

Socio-economic disadvantage affects children's language development in several ways. Many children from low-income households experience a reduced vocabulary or 'word gap' (Golinkoff *et al.,* 2019) and are more likely to enter school with limited lexical knowledge (Hoff, 2014; Huttenlocher *et al.,* 2010). Other children exhibit a 'grammar gap' (Pace *et al.,* 2017), with grammatical skills below age-expected levels. These disparities have important long-term consequences: pre-school language development has been found to be an important determinant of school achievements (Durham et al., 2007), placing early years language skills as a key mechanism for long-term educational outcomes.

Research strongly supports the value of adult–child interactions for promoting oral language development (Rowe, 2022; Rowe and Snow, 2020). Interactions that emphasise talking *with* children rather than *to* them have been shown to increase linguistic abilities. Small-group activities in supportive early years environments are particularly effective (Hassinger-Das *et al.,* 2017; Morra Pellegrino and Scopesi, 1990), especially when facilitated by trained staff who can scaffold children's developing language skills (Dockrell *et al.,* 2017). However, research indicates that many early years settings in disadvantaged areas do not routinely offer activities that promote oral language, such as interactive book reading and structured conversation, and staff often lack training and confidence in supporting language development (Dobinson *et al.,* 2022; Wright *et al.,* 2020). These challenges highlight the need for evidence-based programmes designed to strengthen early oral language support.

Talking Time© is an intervention designed to improve the oral language skills of children aged three to five years. It is designed for settings serving a high proportion of children at risk of language delay but delivered universally, so that all children participate. The programme provides early years practitioners with high-quality professional development and a structured set of small-group activities. The theoretical foundations of the programme draw on evidence that engaging and well-scaffolded small-group interactions offer children opportunities to hear, rehearse, and use language in context (Romeo *et al.,* 2018; Wasik, 2008). Multi-turn conversations are central to the approach, enabling children to build narrative understanding, vocabulary knowledge, and grammatical skills.

The programme was originally developed at University College London (UCL) Institute of Education (Dockrell *et al.,* 2010) and later expanded by the University of Oxford to include a more comprehensive manual (offering scaffolding via flexible plans and prompts) and a structured professional development programme. An initial evaluation of the programme (Dockrell *et al.,* 2010), conducted in three London nurseries in a highly deprived borough, employed a quasi-experimental design, and found significant improvements in vocabulary, oral comprehension, and sentence repetition, though not in narrative skills. The promising findings suggested that Talking Time© could enhance children's oral language abilities but also highlighted the need for a larger, more rigorous experimental study to examine specific pathways of impact.

A subsequent randomised controlled trial (RCT) involving 40 settings, funded by the Nuffield Foundation in 2019/2020, was halted due to the COVID-19 pandemic. A retrial in 2021/2022 with 36 settings (Dockrell *et al.,* 2023) assessed impacts on expressive vocabulary, grammar, comprehension, oral narrative skills, adult–child interaction quality, and parent/carer-

reported language and behaviour. Results showed a significant positive effect on expressive vocabulary, specifically on words embedded in the activities. Trends towards improvements in sentence repetition were observed, though not statistically significant. The trial was again disrupted by absences caused by the COVID-19 pandemic, which limited implementation fidelity and reduced statistical power. The evaluation concluded that further research was necessary to test the intervention under stable conditions, with a larger sample, a longer implementation period, and an examination of how variations in exposure relate to outcomes.

Given these challenges, the Education Endowment Foundation (EEF) agreed to fund an efficacy trial to generate robust evidence on the impact of Talking Time© in disadvantaged early years settings. This new study expands the programme's scope to include both maintained and private, voluntary, and independent (PVI) settings, thereby increasing the generalisability and policy relevance of the findings. The current evaluation also addresses limitations of past trials by incorporating a larger sample size, a longer time frame between baseline and follow-up, and a qualitative follow-up six months post-intervention to gain some insights into long-term changes in practitioner behaviour.

The trial was co-funded by the EEF and four Early Years Stronger Practice Hubs (SPHs): Liverpool City Region and Beyond; A Brighter Start; REACHout; and Thrive Together. In addition to providing funding, these four SPHs, alongside St Edmund's Early Years SPH, supported the Talking Time© delivery team with recruitment to the project. Established by the Department for Education (DfE) in 2022 as part of the Early Years Education Recovery (EYER) programme, SPHs help settings adopt evidence-based practice and provide access to training, expert support, and professional networks. Each participating hub covers a substantial geographical area and provides additional infrastructure to support programme delivery.

The current evaluation uses an RCT to assess the programme's effect on children's oral language development[1] at age three to four years. The evaluation estimates quantitative effects on key language outcomes, while the implementation and process evaluation (IPE) will explore fidelity, mechanisms of change, and practitioners' perceptions of the programme's effectiveness. The study will also include a cost evaluation to assess value for money. An RCT is the most rigorous design available and is essential for determining whether Talking Time© leads to causal improvements in children's language outcomes.

Intervention

Intervention name

Talking Time©.

Why (theory and rationale)

Talking Time© is an oral language development programme for children aged three to five years, which in the current trial is restricted to children aged three to four years attending nurseries. It supports early years practitioners to deliver engaging and structured small-group activities designed to enhance children's oral language skills through evidence-based, high-quality adult–child interactions for all children.

Children from disadvantaged socio-economic backgrounds experience poorer outcomes in standardised assessments of language (Law *et al.,* 2018; Nelson *et al.,* 2011), and early oral language ability is strongly associated with later academic and socio-emotional outcomes. The rationale for the intervention is that all children in early years settings in England need rich and effective English language exposure, particularly those from areas of social disadvantage or those who are dual language learners, which is often lacking in early years settings.

Although Talking Time© is universal and delivered to all children in the target age range within a participating setting, it is expected to have specific benefits for children who are disadvantaged, and for children with EAL. The programme is

[1] Throughout this report, references to children's language skills, language development and language outcomes relate to proficiency in English, as all outcome measures were administered in English. Findings for children with EAL should not be interpreted as reflecting their overall language ability or progress in languages spoken at home

universal at the point of delivery, although it is primarily directed towards settings with higher proportions of disadvantaged children.

The programme aims to enhance the quality of language-supporting practice in early years settings by providing practitioners with high-quality professional development and a structured, evidence-based set of flexible activities. A central focus is supporting educators to use interactive, responsive strategies that promote children's oral language development in conversational contexts. Because it builds staff expertise rather than relying solely on fixed (manualised) programme resources, Talking Time© supports sustainable improvements in language provision.

What (materials)

All participating settings receive:

- a manual with flexible plans and conversation prompts;
- five picture books;
- a starter pack of hexagon photos; and
- a 'Talking Time©' Ted to support the Hexagons activity.[2]

These materials provide prompts and structure (e.g. scaffolding conversations, modelling strategies).

Who (recipients)

Talking Time© is designed for all children aged three to five years. Children younger than three years were not included in the outcome measurement but were not excluded from delivery sessions if they were in rooms where the intervention was implemented. Delivery took place in both publicly funded schools (these include state-maintained nurseries and school-based nurseries that are attached and managed by a primary school) and PVI nurseries; childminder settings were excluded. At the setting level, the programme was targeted, with recruitment prioritised toward settings with higher levels of disadvantage.

The intervention

Talking Time© consists of two integrated components: a structured professional development programme for practitioners'; and a flexible intervention programme for children.

Programme for children

Children participate in two 15-minute structured oral language sessions per week, for 20 weeks, in small groups of up to five children of mixed language ability. One week includes three sessions, rather than two sessions, where the total number of sessions delivered is therefore 41. The programme includes three activity types, all of which emphasise extended conversations and responsive interactions:

- **Story Conversations.** Shared storytelling and conversation using illustrations as prompts.
- **Word Play.** Vocabulary games and guided role-play to develop vocabulary breadth and depth.
- **Hexagons.** Narrative discussion and retelling using photographs of contexts and situations likely to be familiar to the children.

Professional development component

Each setting identifies two Talking Time© leads, who deliver most of the sessions. The decision to appoint two leads was made for pragmatic reasons, and under normal circumstances all nursery staff would be trained through the programme.

[2] Children were given a soft toy Teddy and encouraged to tell it their hexagon story.

A team of Talking Time© trainer-mentors deliver the professional development component and supports implementation across settings.

Each setting is assigned a mentor who conducts the orientation visit, delivers the twilight sessions, and provides in-class and online mentoring to the Talking Time© leads in each setting. Settings designate two Talking Time© leads (typically a teacher and a teaching assistant or equivalent in PVI settings). Leads are responsible for delivering the majority of Talking Time© sessions and are the main point of contact in the setting for the programme. Larger settings may train additional leads. Mentors also maintain contact through weekly emails and encourage ongoing staff reflection.

Talking Time© leads are supported to:

- introduce and embed the three activities of the children's programme;
- use evidence-based language-supporting techniques during sessions; and
- tailor activities to children's specific language needs.

The Talking Time© manual provides initial scaffolding for high-quality delivery, with flexible prompts and plans based on the 'starter' resources provided (picture books, Hexagons photos). However, the overarching goal—supported by the professional development is for practitioners to gradually reduce reliance on programme materials and embed language-supporting activities, strategies, and interactions into everyday practice. The professional development package for practitioners includes:

- Initial welcome visit, including contact with setting leadership.
- Three twilight training sessions (for all staff delivering the programme), introducing the Talking Time© activities, the evidence-based language-supporting strategies, and the professional development approach.
- Four in-class mentoring sessions for the leads which aim to build confidence and expertise in leading Talking Time© sessions, using tailored language-supporting interactions to enhance each child's oracy skills, and reflecting on practice.
- Three online mentoring sessions for each setting's pedagogical lead (in these sessions, practitioners deliver a session or share the recording of a session and discuss it with the trainer-mentor). The sessions also support long-term leadership, planning, and adaptation.
- Weekly individual practitioner reflections, supported by the 'rehearsing, noticing, analysing, and refining' tool.
- Weekly collaboration activities, such as joint planning, discussion, observation, and reflection.

The professional development component is designed to motivate, build knowledge, develop skills, and support embedding of language-supporting practice. It is based on the COM-B (Capability, Opportunity, Motivation, and Behaviour) model of behavioural change (Michie *et al.,* 2011), as reflected in and tested by a recent meta-analysis on the characteristics of effective professional development (Sims *et al.,* 2021) It is underpinned by a framework of six evidence-based language-learning strategies. Throughout the professional development, this framework is used as a reflective tool to support educators in rehearsing, noticing, analysing, and refining their practice (Mathers *et al.,* 2022), helping them move from intuitive to intentional practice. In the final phase of the programme, practitioners are expected to plan their own activities using materials of their choice. The long-term aim is for educators to embed language-supporting activities and interactions throughout daily practice in ways that suit their children and contexts.

Who (provider)

Talking Time© was initially developed by the Institute of Education at UCL (Dockrell *et al.,* 2010). The University of Oxford subsequently created a more comprehensive manual (offering scaffolding via flexible plans and prompts) and developed the structured professional development programme.

How (format and mode of delivery)

Professional development is delivered through a mix of:

- in-person activities (initial visit and in-class mentoring); and
- online components (twilight training and online mentoring sessions).

Programme sessions for children are delivered in small groups of up to five, with children grouped by setting staff into mixed language-level groups. Group membership is typically stable but can be adjusted if children's needs change. Sessions are interactive and encourage active participation.

Where (location)

Children received Talking Time© within their assigned group of five children in their usual early years class as part of everyday activities. Practitioners were encouraged to deliver the programme in a quiet space away from distractions when possible. Practitioner professional development took place both in settings (for face-to-face components) and online.

The evaluation was co-funded by SPHs and took place across five hub regions:

- 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 SPH (Yorkshire and the Humber).

When and how much (dosage)

The programme runs for 20 weeks, with two 15-minute small-group sessions per week. Programme delivery occurred from late 2024 to summer 2025 on a three-week rolling basis to support practical implementation.

Settings were encouraged to continue Talking Time© beyond the formal 20-week period. The final mentoring session focused on planning for sustainability, encouraging settings to sustain Talking Time© with future cohorts, using either the supplied resources or their own materials.

Tailoring (adaptation)

Flexible adaptation is a core principle of Talking Time©. Although manualised, the programme is intentionally designed to be adaptable to local context, practitioner judgement, and children's language needs. The professional development supports practitioners to tailor plans, prompts, and interactions appropriately, and to move progressively towards independent planning.

Theory of change of the intervention

The logic model of the intervention is presented in Figure 1. Developed during the set-up phase of the project, it provides a visual representation of the programme's theory of change. Inputs can be grouped into two categories: those directed at staff delivering the Talking Time© programme (forming the professional development package); and those related to the delivery of the programme to children.

Through the professional development programme, practitioners are expected to engage with a range of activities and resources, including the training manual and the underlying strategies intended to enhance their practice. These activities should lead to improvements in language-supporting practice, as well as increases in practitioner confidence, knowledge, and skills. Ultimately, the programme aims to strengthen language-supporting practice across the setting, supporting successful implementation and contributing to sustained improvements over time.

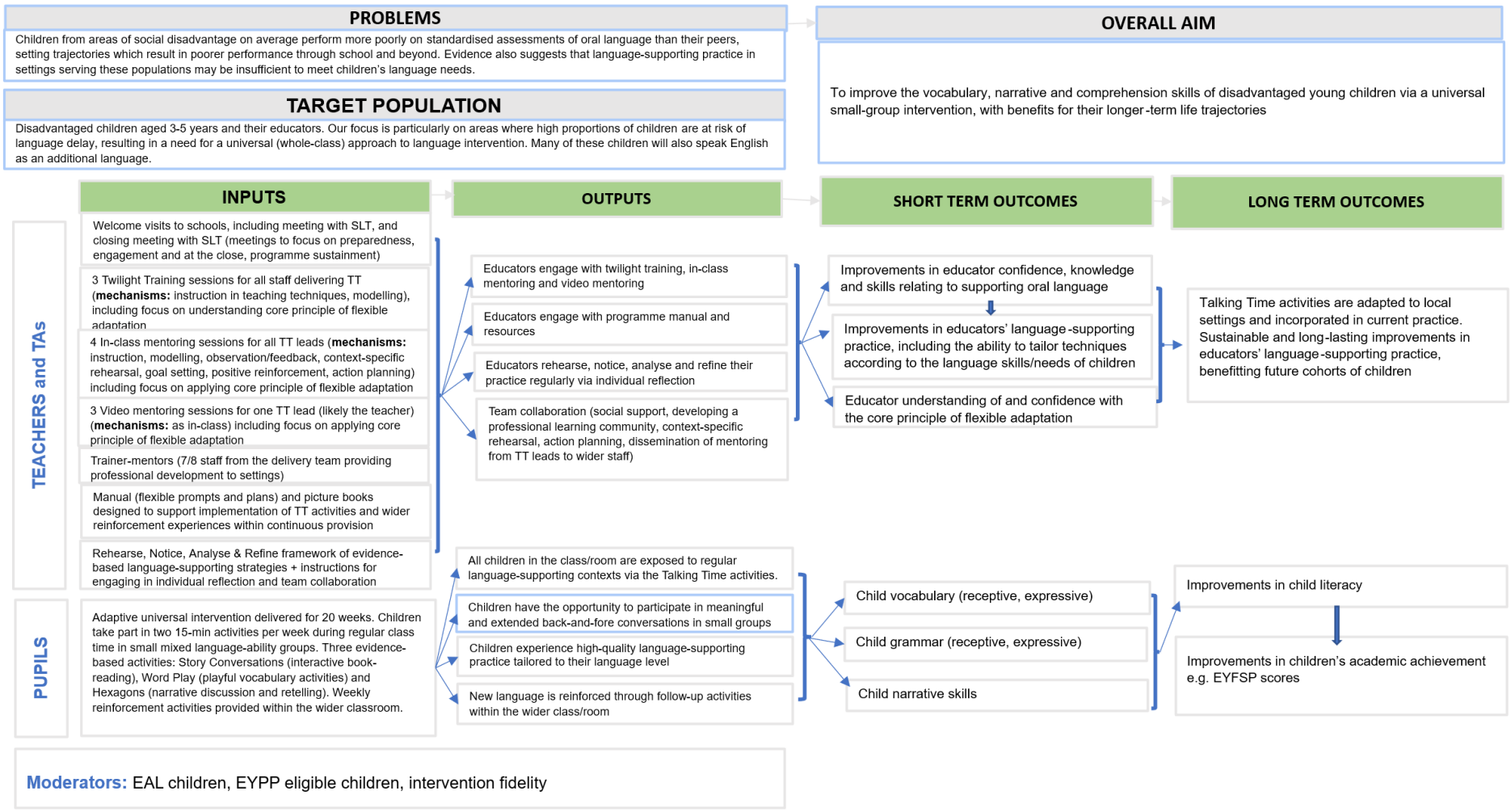
Through the delivery of Talking Time© sessions, the programme aims to provide children with richer language-supporting contexts, including increased opportunities for meaningful small-group conversations and exposure to improved practitioner interactions. As a result of this enhanced exposure, children are expected to make gains in vocabulary, grammar, and narrative skills, which in turn should contribute to broader improvements in literacy skills and later academic attainment.

Because the programme includes a substantial professional development component, with expectations of self-reflection and adaptation, educators are likely to tailor activities to the specific needs and circumstances of their setting. If the programme is successful, these practices are also likely to become embedded in routine provision, generating lasting benefits beyond the formal implementation period.

We identify two moderators that are likely to influence the impact of the intervention: English as an Additional Language (EAL) status and eligibility for Early Years Pupil Premium (EYPP). EYPP eligibility is a broad indicator of low family income. Evidence suggests that children from low-income households growing up in areas of social disadvantage face disproportionate educational delays relative to their more advantaged peers (Law *et al.,* 2017; Nelson *et al.,* 2011). Attending school in a socially and economically deprived area may affect children's language proficiency, and there is evidence that children learning EAL are at greater risk of literacy difficulties in English (August and Shanahan, 2006; Kieffer, 2008). UK data show a strong association between English proficiency and educational outcomes: for example, only 43% of children recorded as 'new to English' achieved a good level of development at the end of Reception, compared with 88% of children recorded as fluent. The gap is larger in early years, although it decreases with age (Strand *et al.,* 2015). Children eligible for EYPP may be more likely to enter nurseries with lower oral language proficiency, while children with EAL may have more limited proficiency in English. Both groups may therefore be more likely to improve their skills as a result of the intervention.

The effectiveness of the intervention is also likely to depend on compliance and fidelity, namely, the extent to which the programme is implemented as intended. Children's attendance at Talking Time© sessions provides one indicator of compliance. However, successful implementation also requires the active engagement of setting leads and staff, which is influenced by the quality of the programme materials and training, as well as contextual factors such as competing work demands, staff capacity, and the availability of functional infrastructure.

Figure 1: Logic model of Talking Time©



Issues that emerged during implementation

Recruiting disadvantaged settings for the intervention proved challenging. Although recruitment targeted more disadvantaged settings, no formal prioritisation based on disadvantage was applied, and all settings willing to participate were included in the study. As it became clear that the recruited settings were not, on average, highly disadvantaged, the team decided to partly address this by prioritising the selection of disadvantaged children within each setting. Children eligible for EYPP were therefore, prioritised during the random selection process for assessments in all settings where the eligible and consented children were in excess of 15 (the maximum number of children assessed per setting).

Evaluation objectives

The primary research question (RQ) that the impact evaluation was designed to address was:

- 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 [REV]) and of information and grammar abilities (Renfrew Action Picture Test [RAPT])?

The following sub-questions of RQ1 were also explored:

- RQ1a. What is the impact of Talking Time© on oral language skills of disadvantaged children that are eligible for EYPP?
- RQ1b. What is the impact of Talking Time© on oral language skills of children with EAL?

The secondary RQ of this project were:

- RQ2. What is the impact of Talking Time© on different aspects of children's oral language skills as measured by the subtests of REV, RAPT, and the sentence repetition assessment from the Grammar and Phonology Screening (GAPS) test?

The following sub-questions of this (RQ2) secondary RQ were:

- RQ2a. For EYPP children, what is the impact of Talking Time© on different aspects of oral language skills, as measured by the subtests of REV, RAPT, and the sentence repetition assessment from the GAPS test?
- RQ2b. For children with EAL, what is the impact of Talking Time© on different aspects of oral language skills, as measured by the subtests of REV, RAPT, and the sentence repetition assessment from the GAPS test?

A final RQ concerned the relationship between outcomes and compliance:

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

The IPE RQs focus on fidelity, context and moderators, and perceived impact. Being an efficacy trial, fidelity is a key concern. Additionally, the evaluation examines the validity of the logic model for a potential future effectiveness trial, considering contextual factors and moderators. Finally, an important factor to the success of the intervention are practitioners' attitudes and expectations, we therefore, include questions on perceived impact. The IPE RQs are:

- IPE RQ1: To what extent were the programme's core components implemented as intended?
- IPE RQ2: 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 RQ3: 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 RQ4: 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?
- IPE RQ5: What are the potential barriers and facilitators to the delivery of the programme? Do these vary by setting type?
- IPE RQ6: 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 RQ7: 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 RQ8: 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?
- IPE RQ9: Do practitioners perceive that children are engaging positively with the sessions? Do practitioners perceive any variation in engagement by children's characteristics, for example, EAL, additional learning needs?
- IPE RQ10: 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 RQ11: 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 RQ12: What 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, for example, 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?
- IPE RQ14: 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?

Ethics and trial registration

The RCT was registered with the International Standard Randomized Controlled Trial Number (ISRCTN) UK Clinical Study Registry with number ISRCTN84530173.[3]

[3] Protocol and statistical plan of the study can be found at the EEF website at the following links: https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/talking_time_-_evaluation_protocol_-_v.2.0.0.pdf?v=1766046073 and https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/talking_time_-_statistical_analysis_plan_-_v.1.0.pdf?v=1766046073

The study design was approved by the National Institute of Economic and Social Research (NIESR) ethics committee after careful consideration and discussion of parent/carer consent, particularly the accessibility of documentation for parents/carers who have EAL, and the need for opt-in consent for participation in the study because of the anticipated video and audio-recordings. The delivery team supplied translated versions of the documentation in a range of home languages for parents/carers with EAL. Regarding opt-in consent, it was made clear on the parent/carer information form that children would not be video- or audio-recorded without explicit parent/carer permission. In addition, since children's assessments were audio-recorded to enable monitoring and verification of scoring quality, it was decided that formal consent would be required from all parents/carers of participating children.

An information sheet for settings provided an overview of the project (see Appendix C), and interested settings were asked to sign a Memorandum of Understanding (MoU) (see Appendix C) outlining in detail the participation requirements and the roles of all parties.

Settings subsequently distributed information sheets to parents/carers of eligible children, explaining the evaluation and how parents/carers could opt-in to the study. The information pack included opt-in consent forms allowing parents/carers to provide permission for their child's participation, which included an authorisation to the audio recording of the baseline and endline assessments. In addition, parents/carers retained the right to withdraw their child from the trial at any point without consequence, and data relating to that child were not used thereafter.

Only children whose parents/carers consented to their participation in the study were included in the evaluation and were assessed at baseline and follow-up (approximately 83% of children). Children were not consulted directly because of their young age; however, assessors remained alert to any signs of boredom or distress during assessments and released children to other activities if such signs were observed.

Practitioners were given an information sheet and asked to provide informed consent to participate in the project. Verbal information was provided prior at each research encounter, and ethical consent was confirmed. This verbal information made clear the purpose of the research, how data would be used, when data would be deleted, and how participants could withdraw their data. Privacy notices detailing the data being collected were also provided to settings, staff, and parents/carers.

Data protection

We recognise the utmost importance of data protection and were committed throughout the Talking Time© evaluation to complying with the Data Protection Act 2018 and the UK General Data Protection Regulation (GDPR). The evaluation required the collection and processing of personal data relating to participating children and staff in early years settings, and all processing was carried out strictly for the purpose of conducting a rigorous and independent evaluation of the Talking Time© programme (privacy notices for parents/carers and practitioners can be found in Appendix C).

The legal basis for processing personal data for the evaluation was GDPR Article 6(1)(f), legitimate interests (GDPR, 2016). NIESR determined that it has a legitimate interest in processing this data in order to fulfil its organisational purpose as an independent research institute and to meet its contractual obligation to deliver the Talking Time© evaluation. The evaluation sought to generate evidence that can improve early years language support, including for children with EAL, which represents a clear societal benefit. The processing was necessary to meet these aims, and the evaluation team took steps throughout to minimise any risks or impacts for participants by limiting the data collected to what was essential, ensuring secure data handling, and providing clear information to all those involved about their rights.

The evaluation also required the processing of special category data, specifically information relating to gender and to languages spoken in the home, which could indirectly indicate ethnicity. These data were processed under GDPR Article 9(2)(j), which permits processing for scientific research purposes (GDPR, 2016), supported by the relevant provisions of the Data Protection Act 2018. This legal basis was appropriate because the evaluation constituted research carried out in the public interest, and information on home language was essential to evaluate whether the programme was effective for

children with EAL. All special category data were processed with safeguards ensuring that findings would be reported only in aggregated, non-identifiable formats and that the data would not be used to make decisions about individual children.

To ensure full GDPR compliance, the evaluation team undertook a Data Protection Impact Assessment (DPIA), which was reviewed by NIESR's data protection officer and kept under review throughout the project. A DPIA was also undertook by the delivery team. Settings and parents/carers were provided with detailed information sheets and privacy notices, produced jointly by the evaluation team, delivery team, and the EEF. These documents explained what data would be collected, how it would be used, participants' rights to withdraw from data processing, and how data would be stored and eventually deleted. Settings also signed an MoU that clearly outlined the data protection and participation requirements. For research interviews, participants were asked to provide informed consent to ensure that they understood the aims of the study, the recording and transcription process, and the safeguards for confidentiality and anonymity.

Data protection compliance was further supported through a Data Sharing Agreement between the delivery team, the evaluation team, and Qa Research (the data collection team). This agreement set out which data would be shared, the methods of transfer, and the responsibilities of each organisation. Only authorised staff had access to identifiable data. Where subcontractors or academic partners were involved in tasks such as transcription, this occurred under formal data processing agreements that required secure and confidential handling of the data.

The evaluation team and the delivery team acted as data controllers for personal data collected during the evaluation, while Qa Research served as a data processor for the child assessments. At the conclusion of the project, and once the final evaluation report is completed, the evaluation team will upload the child assessment data to the EEF archive. At that point, the EEF will become the data controller for the archived dataset, with the Fischer Family Trust (FFT) acting as the data processor.

NIESR will delete all personal data associated with the project three years after the final evaluation report is published. These arrangements reflect our commitment to meeting all GDPR obligations relating to data minimisation, secure storage, restricted access, and timely deletion, while ensuring that the evaluation's findings can contribute to improving early years practice.

Project team

Evaluation team (NIESR)

- **Edoardo Masset, Associate Research Director, NIESR.** Was the principal investigator. Responsible for the overall evaluation study. Led the evaluation team and the study design jointly developed by the evaluation team, the delivery team, and the EEF. Held and carried out the analysis of the data, and wrote the final report. Supported co-ordination between the evaluation team, the delivery team, and the data collection team.
- **Anisa Butt, Economist, NIESR.** Led the organisation and management of data collection, liaised with the delivery team, and contributed to data analysis for the impact evaluation and the writing of the final report.
- **Sherelle Davids, Social Researcher, NIESR.** Responsible for conducting interviews with practitioners, setting leads, and trainer-mentors. Also contributed to analysis and writing the reports.
- **Sophie Kinston, Social Researcher, NIESR.** Contributed to conducting the qualitative IPE interviews.
- **Marley Williams, Associate Social Researcher, NIESR.** Supported the analysis of the qualitative interview conducted under the IPE.
- **Jo Fang, Associate Social Researcher, NIESR.** Supported the analysis of the qualitative interview conducted under the IPE.
- **Claudine Bowyer-Crane, NIESR Fellow and Professor at the University of Sheffield**. Advised the evaluation team on the identification, use, and analysis of language skills indicators. Supported data collection and trained the data collection team.

Delivery team (University of Oxford and UCL)

- **Sandra Mathers, Associate Professor, University of Oxford.** Principal investigator for the delivery team. Led the delivery team, providing intellectual and organisational leadership. Led the development of the professional development programme and expanded the Talking Time© training manual, developed the theory of change for the study, contributed to the overall study design, supported liaison between the delivery and evaluation team, and oversaw recruitment and consent processes.
- **Julie Dockrell, Professor of Psychology and Special Needs at UCL, Institute of Education.** Provided intellectual leadership and expertise on child oracy and research design. Supported the evaluation team's understanding of child language development. Developed the original Talking Time© programme and led the first three evaluations of the programme.
- **Liz Hewitt, Research Officer, University of Oxford.** Led on the organisation and management of Talking Time© delivery. Led on recruitment of settings, the gathering of consent from parents/carers via the settings and liaisons with SPHs. Provided input to many aspects of the IPE evaluation and instruments and—alongside Edoardo Masset—supported the co-ordination between the evaluation team, the delivery team, and the data collection team.
- **Wendy Lee, Professional Director, Lingo.** Worked with Sandra Mathers on the development of the expanded manual and professional development programme and led the mentor team.
- **Keeley Dobinson, Research Officer, University of Oxford.** Supported Liz Hewitt in the organisation and management of Talking Time© recruitment and delivery.
- **Katie Morris, Head of Education Assessments, Qa Research.** Oversaw the evaluation's data collection, ensuring the successful delivery of children testing across all participating schools. Led the recruitment, training, and performance management of a 60-strong education assessor team and directed the planning and co-ordination of the full national testing schedule.
- **Rachel Brown, Education Assessment Manager, Qa Research.** Supported the evaluation's data collection by co-ordinating children testing across participating schools. Served as a key liaison between schools and the 60 education assessors, managing day-to-day communication, and overseeing the scheduling and logistics of all testing visits.

Methods

Trial design

Table 3: Trial design

Trial design, including number of arms		Two-arm RCT with random allocation at the setting level
Unit of randomisation		Setting
Stratification variable (s) (if applicable)		Six geographic areas: East of England (North); East of England (East); London; North-West; West Midlands; Yorkshire and the Humber; and setting type (maintained or PVI)
Primary outcome	Variable	Oral language skills
	Measure (instrument, scale, source)	A composite standardised index of expressive vocabulary—REV scored from 0 to 100 (Renfrew, 2023) and of information and grammar—RAPT information scored from 0 to 41, grammar scored from 0 to 39, (Renfrew, 2019)
Secondary outcome(s)	Variable(s)	1. Sentence repetition 2. Expressive vocabulary 3. Information (vocabulary knowledge) 4. Grammar
	Measure(s) (instrument, scale, source)	1. Sentence repetition score component of the GAPS test scored from 0 to 11 (Gardner *et al.*, 2006) 2. REV scored from 0 to 100 (Renfrew, 2023) 3. Information—RAPT scored from 0 to 41, (Renfrew, 2019) 4. Grammar—RAPT scored from 0 to 39, (Renfrew, 2019)
Baseline for primary outcome	Variable	Oral language skills
	Measure (instrument, scale, source)	A composite standardised index of Expressive vocabulary—REV scored from 0 to 100 (Renfrew, 2023) and of information and grammar—RAPT information scored from 0 to 41, grammar scored from 0 to 39, (Renfrew, 2019)
Baseline for secondary outcome(s)	Variable	1. Sentence repetition 2. Expressive vocabulary 3. Information (vocabulary knowledge) 4. Grammar
	Measure (instrument, scale, source)	1. Sentence repetition score component of the GAPS test scored from 0 to 11 (Gardner *et al.*, 2006) 2. REV scored from 0 to 100 (Renfrew, 2023) 3. Information—RAPT scored from 0 to 41, (Renfrew, 2019) 4. Grammar—RAPT scored from 0 to 39, (Renfrew, 2019)

This study is an efficacy trial, designed to test how the Talking Time© intervention works under ideal and controlled conditions. The trial focuses on assessing impact when the programme implementation is closely monitored to maximise the likelihood of detecting an effect. The evaluation employs a two-arm, cluster RCT design, with early years settings serving as the unit of randomisation. Settings allocated to the intervention arm received the full Talking Time© programme, delivered with support and oversight to ensure adherence to its core components. Settings allocated to the control arm do not receive the programme during the trial period and are expected to continue with business as usual practice.

Control settings receive a £1,000 incentive payment at the end of the trial (with the option to use this to purchase a package of Talking Time© in the year following the trial at a reduced cost). All participating settings, regardless of allocation, also received a £400 thank you payment for supporting the research activities (£150 for baseline assessments in November 2024 and £250 for endline assessments in summer 2025). In addition, the EEF contributed towards the cost of staff time for attending intervention-related training in the project group, reimbursed at £7.50 per hour per educator.

The trial aims to establish the impact of the Talking Time© programme on children's oral language development. The primary outcome is a composite index of language skills, comprising expressive vocabulary measured by REV (Renfrew, 2023), and information (vocabulary knowledge) and grammar measured by RAPT (Renfrew, 2019). The secondary outcomes include the individual components of these assessments, namely, expressive vocabulary (Renfrew, 2023), information (vocabulary knowledge) (Renfrew, 2019), and grammar (Renfrew, 2019), as well as sentence repetition from the GAPS test assessment (Gardner *et al.,* 2006).

After analysing the baseline data, we decided to change the primary outcome of the study because the GAPS test assessment, originally designated as the study's primary measure, proved very difficult to administer in nursery settings and resulted in a high proportion of zero or missing scores (over 40%). A large number of zero scores presents substantial challenges to estimation. Zero reflects censored assessment scores, that would need to be either imputed or excluded from the analysis. Neither approach is desirable because imputation risks introducing bias while exclusion reduces the statistical power of the sample. While the decision to change the primary outcome was made exclusively to address these practical concerns, the revised primary outcome also offers a strong theoretical foundation and a good alignment with the theory of change of the intervention. Oral language is a multidimensional construct that cannot be adequately captured by any single assessment and is better captured by a multidimensional index of different abilities.

Participant selection

The study involved early years settings across England and was open to both state-maintained nurseries, schools with nurseries, and PVIs. Settings were eligible only if they were not participating simultaneously in another SPH programme during the 2024/2025 academic year and were not involved in any trial that included the same children or targeted similar outcomes. Settings taking part at the same time in other early years programmes co-founded by SPH and the EEF, such as Early Talk Boost, The One Programme, Early Years Conversation Project, Concept Cat, or Communication Friendly Settings, were therefore excluded.

The delivery team led the recruitment process and focused on five geographic regions: the North-West; Yorkshire and the Humber; West Midlands; East of England; and London, covered by five early years SPHs (Thrive Together, Liverpool City Region and Beyond Early Years, REACHOut, A Brighter Start, and St. Edmund's Early Years). The professional development component of the intervention required the settings to be clustered to some extent in order to support the logistics of face-to-face interactions with setting practitioners. The hubs and the local authorities participating in the programme are listed in Table 4.

Table 4: Early years SPHs 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, and 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, Kensington and Chelsea, Newham, Redbridge, Tower Hamlets, and Waltham Forest
St Edmund's Early Years (Yorkshire and the Humber)	Doncaster, Calderdale, Kirklees, and Wakefield

Recruitment efforts prioritised disadvantaged areas using the Income Deprivation Affecting Children Index (IDACI) score of child deprivation assigned to the postcode of the setting. Initially, 130 settings were recruited across these regions. Following a small number of withdrawals, 123 settings ultimately joined the trial. Table 5 summarises the regional distribution of settings and the proportion of children included in the study per region at the baseline.

Table 5: Number of settings by region

Region	Recruited settings	Proportion of children at baseline
London	22	0.19
North-West (inc. Yorkshire and the Humber)	31	0.25
East of England	32	0.27
West Midland	38	0.29
Total	123	1.00

At the child level, participation relied on parent/carer consent. Settings distributed parent/carer information sheets and consent forms to all families of attending children and recorded the returned forms using a data collection template provided by the evaluators. Settings distributed an average of 23 forms and received an average of 18 signed forms, yielding an average consent return rate of 83%.[4] Only children whose parents/carers provided written consent were included in the baseline assessments.

There were no exclusion criteria relating to Special Educational Needs and Disabilities (SEND) status; however, both settings and assessors reported that children with SEND frequently struggled to complete assessments, and some staff recommended non-participation. As a result, children with SEND are likely under-represented in the sample.

At recruitment stage, only settings with at least ten children aged three to four years were considered eligible for the intervention, and this requirement was specified in the MoUs with participating settings. However, the expected number of eligible children at recruitment differed from the number observed at randomisation, and not all parents/carers provided consent for their children to participate in the study. As a result, at the time of conducting the assessments, some settings had fewer than ten eligible children. We decided to set a minimum of five consented children per setting for the assessments. This decision reflected both practical and statistical considerations: assessments in settings with very few participating children are inefficient due to fixed administration costs; and small cluster sizes (fewer than five observations) can lead to unreliable variance estimation in mixed-effects models (Eldridge *et al.,* 2006). Although this threshold was established, no setting ultimately fell below it.

While many studies by the EEF use a 15-hour-per-week attendance threshold as an eligibility requirement and this requirement was originally included in the MoUs shared with the participating settings, it was not applied in the present study. As with other whole-class interventions, excluding individual children on the basis of attendance would have been impractical. More importantly, the 15-hour eligibility criterion relied on forecast attendance data provided by settings at baseline which proved to be unreliable. Data collected at the start of the academic year were often incomplete and it became clear that predicted attendance was largely based on guesswork. Many parents/carers had not yet made firm decisions about attendance or had not communicated them to nurseries at the time of data collection. Using these forecasts as an eligibility criterion would have therefore, led to substantial inclusion and exclusion errors.

Finally, applying such a threshold would have reduced the sample's exposure to the intervention. Talking Time© sessions were delivered only twice a week, so overall nursery attendance in hours did not map neatly onto participation in Talking Time©. Excluding children who attended nurseries fewer than 15 hours per week would have removed many who fully participated in Talking Time© sessions but spent fewer hours overall in the nursery. Conversely, it would have retained some children attending nursery more than 15 hours per week who did not participate in any Talking Time© sessions.

Although recruitment targeted disadvantaged areas, baseline data revealed that relatively few consented children were currently eligible for EYPP. In settings with more than 15 consented children (the maximum number assessed per setting), we therefore implemented a random selection process whereby children were selected with probabilities proportional to the historic EYPP distribution in each setting. This approach aimed to ensure that the assessed sample reflected the setting's typical level of disadvantage based on previous years. It should be noted that this procedure was subject to some error, because nurseries did not have full and accurate information on EYPP eligibility for all children at the time of

[4] Based on data reported by 117 of the 123 settings included in the study.

registration. However, it represented the best available information under the circumstances. EYPP eligibility data were subsequently corroborated with settings over the course of the project and were complete and accurate by the end of the trial.

Evaluation preparation study

An evaluation preparation study of the intervention, whose activity were funded by SPHs, was conducted in 19 settings during 2023/2024, prior to the start of the main trial. Within this preparation study, a qualitative study, funded by the EEF, was carried out in five of the 19 settings to gather practitioners' perceptions and feed them back to the delivery team, with the aim of informing the design and implementation of the main trial. The study found that practitioners were enthusiastic about the programme and recognised benefits both for children's educational development and for their own professional development. Some logistic challenges relating, for example, to the timing and content of the sessions, were identified, shared with the delivery team, and subsequently addressed. Settings included in the pilot were not included in the main trial (see Appendix D for the full report of the evaluation preparation study).

In five study settings, the assessments originally specified in the protocol were piloted in April – May 2024 to examine the time required to complete the tests, their ability to capture variability in children's skills, appropriate approaches to code null assessments, and their suitability for use in PVI settings. Analysis of the pilot results led to the removal of one assessment (the Renfrew Bus Story; Renfrew, 2010), the development of a coding tool for missing responses, and revisions to assessment administration procedures.

Outcome measures

Baseline measures

Three assessments, REV, RAPT, and the GAPS test, were administered at both baseline and follow-up. Tests were conducted in one-on-one sessions with children by a team of approximately 40 assessors working across the 123 participating settings. Assessors were locally recruited university students hired, trained, and supervised by Qa Research in collaboration with the evaluation team. All assessments were audio-recorded and Qa Research closely monitored scoring quality to ensure consistency and reliability. Baseline assessments took place primarily before randomisation, ensuring that assessors were blind to treatment allocation, although some mop-up assessments involving less than 5% of children took place during the week after randomisation.

The assessments were selected to capture the broad range of language development outcomes targeted by the intervention. According to the Talking Time© manual, Story Conversations build comprehension and narrative skills through discussion of stories, withy key outcomes including narrative understanding and the use of more complex language. Word Play focuses on expanding vocabulary and deepening understanding of word meanings through structured language activities, with increased vocabulary as the primary outcome. Hexagons reinforce and extend language skills through structured talk, with the key outcome being greater ability to use language flexibly in extended and more abstract conversations.

The outcomes measured by the different test align with these targeted skills. The REV closely captures the effects of World Play as it assesses expressive vocabulary. The RAPT is well suited to capturing the impact of Story Conversations, as it measures both the information conveyed and the grammatical structure of children's responses. It also reflects aspects of the Hexagons, which emphasise extended and structured talk. The GAPS test further captures elements of the Hexagons, particularly in relation to grammar development. An additional assessment, the Renfrew Bus Story (Renfrew, 2010), including its reduced 'Narrative Task' version, was originally planned but was dropped following pilot testing, as most children were unable to respond meaningfully to the test questions.

Primary outcome

The primary outcome of the study is oral language skills, a multidimensional construct encompassing vocabulary, grammar, narrative, and broader expressive language competencies (Massonnié *et al.,* 2022). The Talking Time©

programme is explicitly designed to improve children's oral language across these domains (except phonology) through three structured components:

- **Story Conversations**. Involves joint storytelling 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.
- **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.

Together, these activities are expected to influence multiple aspects of receptive and expressive language, and therefore a multidimensional indicator provides the strongest theoretical and empirical basis for evaluating programme impact.

To capture these language constructs, the primary outcome was measured using a composite index derived from:

- **Expressive vocabulary.** Assessed through the REV (Renfrew, 2023). The full instrument is used without adaptation (it is a commercial test and cannot be reproduced). Children are shown picture prompts and asked to name words of increasing difficulty. The test evaluates a child's ability to correctly name pictures of words arranged by difficulty level. In our pilot, the test took an average of 7 minutes and 55 seconds to complete. The results are scored from 0 to 100.
- **Expressive language (information—vocabulary knowledge—and grammar).** Assessed through RAPT (Renfrew, 2019). A sample card is provided in Appendix E (commercial restrictions apply and the full test cannot be published). RAPT assesses children's grammar skills by asking them to describe a series of illustrated scenes, incorporating words, verbs, and increasingly complex grammatical structures. The children's grammar outcome provides a proxy for quality of connected speech and is scored between 0 and 39. RAPT also measures the amount of relevant information conveyed by the children and their ability to describe actions and details in the pictures. The information score provides an additional measure of vocabulary knowledge and is scored between 0 and 41. In our pilot, the test took an average of four minutes to complete.

The Renfrew assessments are widely used in applied research in the UK. Formal psychometric evidence, such as score reliability and internal consistency, is not discussed in detail in the published manuals and psychometric profiles in peer-reviewed research is limited. The Renfrew scores are available in percentile form against an age-equivalent sample norm from the UK.

Since there is no theoretical basis to assign weights to vocabulary, grammar, or information components of the Renfrew assessments, at the Statistical Analysis Plan (SAP) stage we employed principal component analysis (PCA) to build an index of latent underlying ability using the baseline data. However, PCA delivered factor weights for the three Renfrew measures that were very similar to each other, and the PCA index was basically equivalent to the arithmetic mean of the three measures. We therefore, decided to apply equal weights to the three measures and to use an arithmetic mean to calculate the composite index.

As already mentioned, the use of the composite index implied a significant change to the protocol where the GAPS test assessment had been identified as the primary outcome. After examining baseline data, we revised this decision because the GAPS test was extremely difficult to administer in nursery settings and produced a high proportion (>40%) of zero or

missing scores.[5] In addition to these practical limitations, a composite index better reflects our theoretical understanding of oral language as a multidimensional construct. The change was justified both empirically and theoretically.

Secondary outcomes

The secondary outcomes assess the individual components of oral language listed below. Each outcome provides a focused measure aligned with specific strands of the logic model.

- **Expressive vocabulary.** From REV (Renfrew, 2023).
- **Grammar.** From RAPT (Renfrew, 2019).
- **Information (vocabulary knowledge).** From RAPT (Renfrew, 2019).
- **Sentence repetition.** From the GAPS test (Gardner *et al.,* 2006).

The GAPS test is validated for children aged three years and four months to six years and six months, and reliably identifies language impairment (van der Lely *et al.,* 2011). It evaluates fundamental grammatical skills across sentences and words. The test can be administered by both professionals and non-professionals, typically taking five to ten minutes to complete. During our pilot, it took an average of 5 minutes and 20 seconds to complete. In our study, we only used the grammar component of the assessment (sentence repetition), because phonology is not targeted by the intervention. The grammar component of the GAPS test includes 11 questions. The child is asked to repeat a series of short sentences after the administrator's prompt, with the support of a visual aid booklet story. The test is scored from 0 to 11. The instrument is publicly available online and reported in Appendix E.

The assessments were administered by university students who were trained by the data collection team with support of the evaluation team. All assessments were audio-recorded for the purpose of monitoring adherence to administration protocols and verifying scoring accuracy. All audio-recordings were individually played to double-check scores. At baseline the administration was fully blinded since randomisation took place after the assessments had been completed (with the exception of a small percentage of children assessed during mop-up sessions). At follow-up assessors remained blinded to treatment assignment and setting staff were instructed not to disclose the assignment status of the setting to the assessors.

The GAPS test has good construct and concurrent validity.[6] It is highly correlated with other established language assessments. It also has good internal consistency reliability (approximately 0.86 for the sentence repetition component used in our study). It also has good diagnostic accuracy, with reported high sensitivity and specificity for identifying children with specific language impairment (van der Lely *et al.,* 2011). It was originally standardised on a UK population of 668 children in the age group three years and four months to six years and six months (Gardner *et al.,* 2006). Raw scores can be compared to an age-specific UK norm.

Sample size

Protocol stage

Sample size calculations were conducted using Optimal Design Version 3.01 using the following assumptions: an equal number of intervention and control settings (total of 130); a two-sided 5% significance level; an 80% statistical power; a pre-test/post-test correlation of 0.40 (drawing on evidence from the EEF studies showing correlations between 0.43 and 0.63

[5] During the assessment pilot it was observed that many children were unable to answer any of the GAPS test item questions, as the task was perceived too difficult, and they frequently lost concentration and attention during administration. It was therefore decided to introduce a stopping rule, whereby the assessment would stop if the child was unwilling or unable to provide a response for the first three items. Many of the zero scores observed on the GAPS test assessment at the baseline resulted from the application of this stopping rule. The stopping rule was only used at baseline and was abandoned at follow-up.

[6] See the EEF measures database for the GAPS test at: **EEF | Grammar and Phonology Screening Test**

using National Pupil Database [NPD] data and between 0.54 and 0.88 in the EEF previous trials); an intracluster correlation coefficient (ICC) of 0.10 (slightly higher than the 0.06 typically observed in early years English outcomes to be on the conservative side); and an average of 20 eligible children per setting reduced to 15 eligible children after allowing for significant 25% attrition. Based on national data and the targeted nature of recruitment, we assumed that 20% of children would be eligible for EYPP, which serves as the study's indicator of socio-economic disadvantage.

Under these assumptions, the expected minimum detectable effect size (MDES) was 0.17 SD for the full sample and 0.29 standard deviation (SD) for the EYPP subgroup. Thus, while the trial was powered to detect impacts for the overall cohort, it was not powered to detect plausible effects for children eligible for the EYPP.

Randomisation stage

At the randomisation stage, we repeated the statistical power calculations specified in the SAP (Masset and Butt, 2025) using the same parametric assumption but updating the number of settings and children based on actual baseline figures. Recruitment took place between January and July 2024, with a short extension to replace early withdrawals. In total, 123 of the 130 settings identified at the protocol stage joined the study. In addition, several settings had smaller than expected number of children while some parents/carers did not provide consent, resulting in a reduction of the average number of assessed children per setting from 15 to 13. At this stage it was also decided to cap the number of children assessed at 15 in the selected settings. In a cluster RCT, there are limited gains from collecting data on large numbers of children within the same cluster. Given budget constraints on the overall sample size, we therefore, prioritised spreading the number of assessed children as evenly as possible across clusters, rather than concentrating large numbers within a small number of settings. When the size of the cluster was larger than 15, we therefore, randomised 15 children for the study, assigning greater probability of selection to children eligible for EYPP as mentioned earlier.

The reduction in the number of settings and in the average cluster size led to only a modest increase in the MDES, which rose to 0.20 SD for the full sample and to 0.36 SD for the EYPP subgroup. The trial therefore remained adequately powered to detect effects in the full sample but continued to be underpowered for detecting effects among EYPP-eligible children.

Analysis stage

Power was re-assessed at the analysis stage using the achieved sample size and the empirically observed ICC and pre-test/post-test correlations. The sample used in the analysis was further affected by missing assessment data: although each child was eligible for four assessments, not all children completed all tests, and the analysis could only include children who completed the same assessment at both baseline and follow-up. Table 10 reports the MDES for the three samples outlined above (protocol stage, randomisation, and analysis stage). The latter includes only the children for whom both baseline and endline assessments were available.

The sample at the analysis stage was considerably smaller than at the protocol stage (an average of 12 children per setting against 15 children at the protocol stage). However, no settings were lost during the trial as a result of missing observations. The ICC in the analysis sample was very similar to the ICC considered at the protocol stage (0.11 against 0.10).[7] However, the pre-test/post-test correlations were more favourable than estimated at the protocol and SAP stages. The pre-test/post-test correlation for the primary outcome was considerably higher than assumed (0.82 compared with 0.40). This parameter offset the reduction in sample size, allowing the achieved MDES for the primary outcome to remain aligned with the protocol stage estimate of 0.17 SD.

These improvements did not sufficiently reduce the MDES for the EYPP subgroup, which remained large (0.36 SD) due to the relatively small size of this sub-sample. As a result, although the study is appropriately powered for detecting effects in the full population, it is not powered to detect impacts for EYPP-eligible children.

[7] ICC were calculated from the data in relation to the primary outcome (the combined Renfrew test index). See Table 34 (coefficient for the Renfrew index, model with strata, follow-up).

Randomisation

The trial was a two-arm, setting-level RCT, with early years settings as the units of randomisation. Randomisation took place during the week of 04 November 2024, one week prior to the intervention start (11 November 2024). A total of 123 settings were randomised: 62 to the intervention group; and 61 to the control group. The evaluation principal investigator generated the random number sequence, implemented the randomisation, and produced the final allocation. Randomisation was performed in Stata v17 (StataCorp LLC, College Station, Texas, USA) and the full randomisation code is available in Appendix F. The allocation list was then shared with the delivery team, which communicated group assignments to settings during the following week.

A stratified randomisation design was used, with 12 strata defined by six geographic regions: London; North-West with Yorkshire and the Humber; North-West (Liverpool); East of England—North; East of England—East; and West Midlands, crossed with PVI status. Settings are therefore, divided into two groups, PVIs and non-PVIs. Non-PVI settings include maintained nurseries and school-based nurseries. The two are similar in some respects as they are both part of the state-funded education system and led by qualified teachers, but they also differ in terms of governance and internal administration. Ultimately, they were considered together for pragmatic reasons. They are more similar to each other than to PVI settings and given the number of clusters we needed to limit the number of strata. Within each stratum, allocation followed an implicit block structure: after randomly assigning the first setting (based on a uniform draw of 0–1), subsequent settings within that stratum were alternated between treatment arms using the sorted random sequence.

Because some strata were very small, introducing greater risk of imbalance between intervention and control observations, we used a restricted randomisation procedure following Hayes and Moulton (2017). The aim was to select an allocation that minimised imbalance in factors expected to predict the primary outcome. The restricted procedure was implemented using the following algorithm (full reproducible Stata syntax and execution logs are provided in Appendix G):

- For each random allocation, we regressed the treatment assignment indicator on three pre-specified setting-level characteristics: the proportion of children eligible for EYPP; the proportion of children with EAL; and the IDACI deprivation index score.
- We repeated the randomisation procedure 10,000 times.
- We recorded the R^2 from each regression as a summary measure of imbalance (large R^2 meaning relatively larger imbalance).
- We selected the allocation with the lowest R^2, indicating the allocation in which treatment status was most independent of the selected characteristics.

Table 6 shows the number of settings per strata and treatment arm. Note that one of the strata had only one setting and was therefore merged to another stratum (for the purpose of randomisation, a PVI in the East of England (East) was included in the PVI London stratum). Table 6 also shows the number of children per strata and arm using the sample of children that were assessed (with any of the assessments) at baseline and randomised.

The evaluation team was blinded to treatment allocation when preparing the analysis syntax, data-cleaning procedures, and the pre-analysis plan. Setting and strata labels were masked during data preparation. Unblinding occurred only after the final syntax had been locked and approved.

Table 6: Distribution of settings and analysis sample, by strata

Strata	Settings			Children	
	All settings	Intervention	Control	Intervention	Control
State-maintained: North-West and Yorkshire and the Humber	9	5	4	75	64
State-maintained: North-West (Liverpool)	13	6	7	80	104
State-maintained: West Midlands	32	16	16	222	216
State-maintained: East of England (North)	16	8	8	110	105
State-maintained: East of England (East)	10	5	5	77	72
State-maintained: London	17	9	8	147	115
PVI: North-West and Yorkshire and the Humber	6	3	3	43	35
PVI: North-West (Liverpool)	3	1	2	16	23
PVI: West Midlands	6	3	3	30	29
PVI: East of England (North)	5	3	2	37	34
PVI: London	6	3	3	35	36

Statistical analysis

Primary analysis

To address the primary RQ (RQ1), we estimate the impact of the intervention on children's oral language skills using a multilevel analysis of covariance (ANCOVA) model. The primary outcome is a composite index of the Renfrew tests. For both baseline and follow-up, we use the simple arithmetic mean of the three Renfrew assessments scores. The three assessments use different scoring scales. To avoid that the index reflects these scale differences we standardise the individual scores and we take the mean of the standardised scores. Our index is:

$$I = \frac{z_{i1} + z_{i2} + z_{i3}}{3}$$

where z_{ik} are the three Renfrew assessments standardised scores using the baseline mean and the baseline SD across the intervention and control groups in the following way:

$$z_{ik} = \frac{score_{ik} - \overline{score_k}}{SD_k^{pooled}}$$

The intervention effect is identified by comparing post-intervention scores between the treatment and control groups while adjusting for pre-intervention scores. Our main specification regresses endline scores on baseline scores, the treatment indicator, and the stratification variables used at randomisation. In the section "Additional analyses and robustness checks" we also report a fully 'saturated' ANCOVA specification that includes additional child- and setting-level baseline characteristics.

As outlined in the SAP (Masset and Butt, 2025), we estimate the effect size using a multilevel random intercept model. This approach is preferred to the fixed-effects ordinary least squares (OLS) model described in the protocol because the multilevel model accounts for unobserved heterogeneity between settings, whereas OLS does not.

In both the primary and secondary outcome analyses, we estimate the intention-to-treat (ITT) effect of the intervention, namely, the average impact of the intervention on the population, regardless of programme compliance or treatment dosage. All children assessed both at baseline and follow-up are included in the analysis, whether or not they actively participated in the intervention or dropped out during implementation.

The multilevel ANCOVA model is:

$$y_{uit} = \beta_1 y_{uit_{-1}} + \beta_2 T_{ui} + \sum_{s=1}^{11} \gamma_s S_{si} + \alpha_u + \varepsilon_{iu}$$

- y_{uit} is the test score of child *i* in setting *u* at the endline *t;*
- β_1 is the correlation between the scores for the same child before (*t*-1) and after the intervention (*t*);
- T_{ui} is treatment status for child i in setting u (T=1 if the child is in the intervention group);
- β_2 is the estimated impact of the intervention;
- S_{si} are indicators for the strata used at randomisation;
- α_u are setting-level random-effects; and
- ε_{iu} is a child-level error term.

Secondary analysis

In the secondary outcome analysis, we estimate the impact of the intervention on the GAPS test score and on the individual Renfrew test scores (REV and RAPT). In all cases, we use raw scores rather than standardised scores. The analysis of secondary outcomes follows the same approach used for the primary outcome, based on the multilevel random intercept ANCOVA model:

$$y_{uit} = \beta_1 y_{uit_{-1}} + \beta_2 T_{ui} + \sum_{s=1}^{11} \gamma_s S_{si} + \alpha_u + \varepsilon_{iu}$$

- y_{uit} is the test score of child i in setting u at the endline *t*;
- β_1 is the correlation between the scores for the same child before (*t*-1) and after the intervention (*t*);
- T_{ui} is treatment status for child i in setting u (T=1 if the child is in the intervention group);
- β_2 is the estimated impact of the intervention;
- S_{si} are indicators for the strata used at randomisation;
- α_u are setting-level random-effects; and
- ε_{iu} is a child-level error term.

We estimate the model separately for the expressive vocabulary score, for the information (vocabulary knowledge) and grammar RAPT scores, and for the GAPS test measure. Since oral language is a multidimensional construct, and different components of the Talking Time© programme target different language skills, this analysis will help identify which specific domains were most strongly affected by the intervention. Examining each test individually will therefore provide insight into which activities within the programme were most effective in strengthening particular aspects of children's oral language development.

Subgroup analysis

Children from more disadvantaged backgrounds perform, on average, more poorly on standardised assessments of oral language (Law *et al.,* 2018; Nelson *et al.,* 2011). Attending school in socially or economically deprived neighbourhoods may further constrain language development, and there is also evidence that children learning EAL face an elevated risk of later literacy difficulties in English (August and Shanahan, 2006; Kieffer, 2008).

We consider the impact of Talking Time© on subgroups of EYPP children and children with EAL. To analyse subgroup effects, we use two complementary approaches: i) estimating the intervention effect separately for children within and outside the subgroup; and ii) estimating a single model for the full sample that includes an interaction between subgroup status and treatment. In both cases, we apply the same model specification as in the primary analysis, namely, a multilevel random intercept ANCOVA model.

For each subgroup, we estimate a single model using the full sample and including both subgroup status and its interaction with treatment status:

$$y_{uit} = \beta_1 y_{uit_{-1}} + \beta_2 T_{ui} + \sum_{s=1}^{10} \gamma_s S_{si} + \partial_1 subg_i + \partial_2 subg_i * T_{ui} + \alpha_u + \varepsilon_{iu}$$

where:

- y_{uit} is the test score of child *i* in setting *u* at the endline *t;*
- β_1 is the correlation between the scores for the same child before (*t*-1) and after the intervention (*t*);
- T_{ui} is treatment status for child i in setting *u* (*T=1* if the child is in the intervention group);
- β_2 is the estimated impact of the intervention;
- S_{si} are indicators for the strata used at randomisation;
- α_u are setting-level random-effects;
- ε_{iu} is a child-level error term;
- $subg_i$=1 if the child belongs to the subgroup of interest (EYPP or EAL); and
- ∂_2 captures whether the intervention effect differs for that subgroup.

A statistically significant $\delta_2$indicates that the effect of Talking Time© differs for the subgroup relative to other children. The implied intervention effect for subgroup members is: β_2+∂_2.

We conduct such subgroup analysis for all four study assessments. Since we are conducting two subgroup analyses for four different assessments, there is a risk of obtaining statistically significant results by chance. To account for multiple testing of hypotheses, we set-up an algorithm that uses a False Discovery Rate (FDR) adjustment (Efron and Hastie, 2016), specifically the Benjamini–Hochberg procedure. Under this approach we first order the p-values of our tests from smallest to largest, indexing them by i. For each ordered p-value we compute the threshold value:

$$tr_i = \frac{i}{N} q$$

where i is the rank of the p-value, N is the total number of tests (here, N=8) and q is the targeted FDR level (commonly set at 0.10). We then compare each observed p-value to its corresponding threshold value (tr) and identify the largest I for which $p_i \leq tr_i$. All tests with p-values less then or equal to this threshold are considered relevant.

Analysis in the presence of non-compliance

The ITT analysis estimates the effect of assignment to the intervention, regardless of whether children actually received the Talking Time© sessions. To estimate the effect of the intervention on children who received it, we calculate the Complier Average Causal Effect (CACE), also known as the Local Average Treatment Effect (LATE). This represents the impact of the intervention on children who complied with their assigned treatment status (i.e. compliers).

We define compliance at the child level, based on the number of Talking Time© sessions each child attended. Attendance was recorded by programme practitioners during each session and collated by the delivery team and the evaluation team.

Since there is no theoretical or empirical justification for selecting a fixed attendance threshold, our main measure of compliance is continuous, defined as the total number of Talking Time© sessions attended (range: 1–41).

In addition, and for exploratory purposes, we construct a threshold-based measure of compliance. Based on guidance from the project delivery team that a meaningful effect requires attendance at ≥50% of sessions, we defined compliance using thresholds ranging from 50% to 100% attendance. This approach should not be interpreted as estimating a dose–response function. CACE estimates intervention effects that are local to the selected sample. As the threshold changes the sample also changes selecting characteristics that are correlated with response to the treatment. This approach does not represent the population response to increasing exposure to the intervention. Despite this limitation, the CACE effect by threshold attendance allows us to explore whether CACE estimates are stable (monotonic) across different plausible definitions of compliance and to examine sample heterogeneity. The approach also avoids selecting an arbitrary level of compliance for the analysis.

The approach described above differs slightly from that set out in the SAP (Masset and Butt, 2025), which proposed analysing compliance for: i) children attending all Talking Time© sessions; and ii) children attending at least 50% of the sessions. The change was made because there was no strong theoretical justification for the thresholds specified in the SAP (Masset and Butt, 2025), and they did not align well with the empirical distribution of session attendance observed in the data. In practice, attendance was higher than anticipated, such that a 50% threshold was not informative, and it was distributed along a continuum, supporting a continuous rather than dichotomous approach to analysing compliance. The change was reviewed and approved by the EEF.

CACE estimation relies on two assumptions (Imbens and Rubin, 2015). The first assumption is unconfounded treatment assignment. This assumption is met by design. Since the intervention was randomised at the cluster level, we can estimate both the ITT effect and the impact of assignment on intervention receipt without confounding. The second assumption is the exclusion restriction. We assume that assignment to treatment does not affect outcomes for non-compliers. Although double-blinding is not feasible in this context, the assumption is plausible in a cluster-randomised design. Nonetheless, we acknowledge the possibility that practitioners in control settings may adjust their behaviour in response to study participation.

Under these assumptions, CACE can be expressed as:

$$CACE = \frac{ITT}{P(T)}$$

The CACE effect is the ratio of the impact of the intervention assignment on the outcome (ITT) to the impact of assignment on intervention receipt. ITT is the effect size of the intervention estimated in the primary analysis, and the probability of receiving the intervention (i.e. the proportion of compliers) can be calculated from the data.

We estimate CACE using a two-step instrumental variable regression approach. First, we regress compliance on treatment assignment and baseline covariates:

$$comp_{ui} = \beta_1 T_{ui} + \sum_{k=1}^{n} \beta_k X_{kit_{-1}} + \varepsilon_i$$

- $comp_{ui}$ is child attendance of Talking Time© sessions in setting *u*;
- T_{ui} is treatment assignment (*T*=intervention);
- $X_{kit_{-1}}$ are *k* child-level (age in months, gender, EYPP status, EAL status) and setting-level variables (IDACI index) baseline covariates; and
- β_2 is the estimated impact of assignment on compliance.

We obtain predicted values $\widehat{comp}_{ui}$ from this first-stage model. Second, we then estimate the model for the primary outcome using predicted compliance from the first-stage model:

$$y_{uit} = \beta_1 y_{uit_{-1}} + \beta_2 \widehat{comp}_{ui} + \sum_{k=1}^{n} \beta_k X_{kit_{-1}} + \varepsilon_i$$

where:

- y_{uit} child's endline test score;
- $y_{uit_{-1}}$ is baseline score;
- $X_{kit_{-1}}$ are k child-level (age in months, gender, EYPP status, EAL status) and setting-level variables (IDACI index) baseline covariates;
- $\widehat{comp}_{ui}$ is predicted compliance; and
- β_2 is the estimated CACE effect.

Missing data analysis

We define three types of missingness (Rubin, 1976): missing completely at random (MCAR); missing at random (MAR); and missing not at random (MNAR). Data are MCAR, when the missingness is uncorrelated with the outcome. For example, some assessors may be less efficient at administering the tests and complete fewer tests. Data are MAR when there is an association between missingness and observable characteristics, but after conditioning on these characteristics, missingness is random. For example, missingness of post-intervention test scores could be related to EYPP eligibility, but conditional on EYPP (that is within groups of EYPP and non-EYPP children) missingness is random. Data are MNAR when there is an association between missingness and unobservable characteristics that are correlated with the outcome. For example, for some reasons, children in the intervention group that are expecting to perform poorly in the test, may systematically refuse to be tested.

The appropriate analysis depends on which type of missingness may occur. There is no test that can definitively distinguish MCAR from MAR. We therefore explore the patterns of missingness, and we make assumptions based on these results.

Specifically, we take the following steps:

- compute the number and percentage of missing cases at baseline and endline for the outcome variables;
- report and discuss the reasons for missing cases as reported by the assessors;

- compare and test the difference in the proportion of missing cases between intervention and control group at the baseline and endline; and
- estimate a missingness model to identify variables correlated with missingness separately for the intervention and the control group and test the differences in the coefficients after running the logistic regression.

$$log\left(\frac{P(y-1)}{1-P(y-1)}\right) = \beta_0 + \beta_1 T_{ui} + \sum_{k=1}^{n} \beta_k X_{ki}$$

Where:

- *y=1* if the observation is missing and 0 otherwise;
- *P* is the probability that the observation is missing;
- T_{ui} is treatment status; and
- X_{ki} are child-level (gender, age in months, EYPP status, EAL status) and setting-level (IDACI index) baseline covariates.

If missingness exceeds 5% we examine the data to understand its extent and its determinants. We then perform the following tentative classification of missingness:

- **MCAR.** If there are no treatment arm differences and no covariates associations.
- **MAR.** If covariates explain missingness but no arm differences remain.
- **MNAR.** If arm differences persist after controlling for covariates and missingness patterns differ substantially between arms.

If the evidence supports MCAR, we perform the analysis using only complete cases. All observations with missing baseline or endline outcome scores are omitted. If missingness appears to be MAR, we re-estimate the primary ANCOVA model including the covariates associated with missingness. Finally, if the evidence suggests a MNAR mechanism, we estimate a Heckman-type selection model to correct the estimates (Muñoz *et al.,* 2023). In addition, we will conduct the sensitivity analyses recommended by Van Buuren (2018) for non-ignorable models.

Additional analyses and robustness checks

In our robustness analysis, we compare the results of the multilevel random intercept model used in our primary analysis to:

- an OLS model with fixed effects;
- a multilevel random intercept and random slope model; and
- saturated models including baseline covariates.

We estimate the following OLS model with fixed effects and standard errors clustered at the setting level:

$$y_{uit} = \beta_1 y_{uit_{-1}} + \beta_2 T_{ui} + \sum_{s=1}^{11} \gamma_s S_{si} + \varepsilon_{iu}$$

where:

- y_{uit} is the endline test score of child *i* in setting *u*;
- β_1 is the correlation between baseline and endline scores;
- T_{ui} is treatment status;
- S_{si} are the 11 strata indicators used at randomisation;
- β_2 is the estimated impact of the intervention; and
- ε_{iu} is a child-level error term.

We then estimate the following multilevel random intercept and random slope model:

$$y_{uit} = \beta_1 y_{uit_{-1}} + \beta_2 T_{ui} + \sum_{s=1}^{10} \gamma_s S_{si} + \alpha_u + \alpha_u T + \varepsilon_{iu}$$

where the variables and coefficients have the same meaning as those used in the multilevel random intercept model used in the primary analysis, and in addition:

- α_u are setting-level random-effects (random intercepts); and
- $\alpha_u T$ are interactions between the random intercept and treatment status (random slopes).

Unlike the random intercept model, which assumes homogeneous treatment effects across settings, the random slope model allows the treatment effect to vary across settings. This is appropriate if, for example, setting differ considerably in terms of implementation capacity, staff characteristics, or population composition. The random slope model allows the impact of the intervention to vary across settings, which may happen, for example, as a result of differences in implementation or in the characteristics of the population. To assess whether allowing random slopes improves model fit relative to the random intercept model, we conduct a likelihood ratio test.

Finally, we run 'saturated' versions of the models above, which include pre-intervention child-level and setting-level covariates. These covariates are included to improve precision and to increase balance between treatment and control groups. The multilevel random intercept model used in the primary outcome analysis becomes:

$$y_{uit} = \beta_1 y_{uit_{-1}} + \beta_2 T_{ui} + \sum_{s=1}^{10} \gamma_s S_{si} + \sum_{k=1}^{n} \theta_k X_{kit_{-1}} + \alpha_u + \varepsilon_{iu}$$

where $X_{kit_{-1}}$ are child-level baseline covariates: (age in months, gender, EYPP status, and EAL status) and setting-level covariates (IDACI index). We employ the same covariates to extend the OLS fixed effect model and the random slope random coefficient model.

Estimation of effect sizes

We convert the impact estimates obtained using the ANCOVA random intercept multilevel mode into standardised effect sizes to facilitate comparability across studies. In line with the EEF statistical guidance (EEF, 2022), we define the effect size as the adjusted mean difference divided by the pooled unconditional SD:

$$ES = \frac{\left(\underline{Y_t} - \underline{Y_c}\right)_{adjusted}}{sd_{unconditional}}$$

where:

- $\left(\underline{Y_t} - \underline{Y_c}\right)_{adjusted}$ is the ANCOVA-adjusted difference in means between the treatment and the control group (adjusted for pre-test score and the stratification variables); and
- $sd_{unconditional}$ is calculated using the pooled unconditional SD of the treatment and control groups:

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

We calculate CIs using the standard normal approximation:

$$CI_{95\%} = estimate \ \pm \ 1.96 \, se(effect)$$

Estimation of ICC

We report ICC for all pre-intervention and post-intervention assessments at the setting level. Observations within clusters are correlated, and this has implications for both the calculation of standard errors and the statistical power of the study. In a two-level data structure (children nested within settings), the degree of similarity among observations in the same cluster is quantified by the ICC.

$$ICC = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2}$$

where:

- σ_b^2 is the between-cluster variance; and
- σ_w^2 is the within-cluster variance.

The ICC represents the proportion of total variance attributable to differences between clusters. Both variance components are obtained from the multilevel models, making the ICC straightforward to compute.

IPE

Research methods

The IPE was conducted utilising semi-structured interviews and baseline and endline surveys. This generated both qualitative and quantitative data.

Semi-structured interviews

The evaluation team conducted interviews with a sample of practitioners and setting leads from intervention settings across different setting types and regions. Control setting leads and trainer-mentors were also interviewed. Sampling aimed to ensure regional representation, a mix of school-based and PVI settings, and a prioritisation of settings serving high numbers of disadvantaged children and/or those with EAL. Table 7 provides a breakdown of participants and rationale for their inclusion. In total, 40 interviews were completed with intervention practitioners, intervention setting leads, and control setting leads across 23 settings. A further seven interviews were conducted with trainer-mentors, these were all the trainer-mentors on the programme. Interviews took place online, lasting up to one hour.

Practitioners were those responsible for the day-to-day delivery of activity of children, for example, a teacher in a school or a room leader in a PVI setting. Setting leads were those in leadership positions, usually the headteacher or another member

of the senior leadership team (SLT) in schools and the nursery manager in PVIs. Trainer-mentors were responsible for delivering the professional development component of Talking Time©. The Talking Time© intervention was delivered by practitioners or setting leads assigned as Talking Time© leads. In schools, Talking Time© leads were typically a qualified teacher paired with a junior staff member (e.g. teaching assistant), while in PVIs it was often the case that the setting lead (i.e. nursery manager) was one of the Talking Time© leads.

Interview topic guides were produced by the IPE team. These were edited based on feedback from the delivery team and the EEF. The aim of the interviews with practitioners were to address dimensions of the IPE that relate to fidelity, implementation/context (in particular the enablers and challenges to delivering Talking Time©), perceived impact, and programme differentiation. Importantly, as adaptability is a key feature of Talking Time©, interviews with practitioners were a key source of data for understanding what and how principled adaptations were enacted as well as capturing information on adaptations that were not within the remit of the programme. Fidelity, implementation, and perceived impact were the focus of interviews with setting leads and trainer-mentors. Setting leader interviews aimed to understand how existing strategy around professional development and the early years curriculum compared to Talking Time©, while interviews with trainer-mentors were particularly useful for understanding practitioner engagement with professional development sessions. Interviews with control settings addressed programme differentiation as well as monitoring of the control group.

The semi-structured designed of interviews allowed the IPE to gather in-depth data on experiences and perceptions from the point of view of those with different roles in the delivery of Talking Time©. The format of the interviews was both structured and flexible, ensuring the necessary dimensions of the IPE were covered while allowing the opportunity for themes not anticipated to arise. During the course of conducting the first few interviews, the IPE team met to discuss the usefulness of the topic guides. This resulted in the reordering of some questions in order to create a better flow of conversation.

Surveys

The evaluation team conducted baseline and follow-up practitioner surveys to capture information on practices, implementation, and perceived impacts of Talking Time©. At baseline, the survey was administered across participating settings to setting leads in order to establish pre-intervention context. At endline, separate surveys were administered to Talking Time© leads in the project and control settings to reflect differences in programme exposure.

The baseline survey was completed by 115 respondents across 123 participating settings, provide strong evidence on existing practice prior to the implementation of Talking Time©. The survey aimed to understand the context in which the programme was delivered, including general characteristics of settings, staff professional development, and current approaches to supporting children's oral language. The baseline survey data establish a detailed picture of pre-intervention conditions across settings. This information is essential for interpreting changes observed at endline and for distinguishing programme-related effects differences in provision. The results of the baseline survey are presented in the subsection on 'Children and setting characteristics' in the 'Impact evaluation results' section below and are used to provide a baseline picture of key setting characteristics.

At endline, 123 settings were participating in the evaluation. In total, 98 responses were received, 55 from intervention settings and 43 from control settings. The endline survey for project settings focused on experiences of implementing Talking Time© and perceptions of its impact. Questions covered several key areas such as general setting information, children's engagement, and perceived impact of Talking Time© sessions; the usefulness of programme resources, including the handbook, storybooks, and Hexagons sets; professional learning and mentor support; and practical aspects of programme delivery. The survey also explored peer and leadership support, reflections on how Talking Time© differed from previous oral language practices, and intentions to continue using the approach in the future. In addition, respondents were asked about general practices within their setting beyond Talking Time©, providing broader contextual information. The endline survey for control settings was more limited in scope, focusing on general information and existing practices within settings. These questions were designed to capture practice in the absence of Talking Time©, providing an

appropriate comparator for interpreting changes observed in project settings. The results of the endline survey are discussed throughout the IPE section.

Session logs

The delivery team collected Talking Time© session logs from each setting. For each child in the study, settings recorded attendance at each of the 41 Talking Time© sessions over the course of the programme.

Follow-up interviews

A series of interviews with practitioners were undertaken in 2026, several months after the end of the trial, to understand: whether and how Talking Time© continues to be delivered; the extent to which core principles are maintained or adapted; the factors influencing how the programme is embedded over time; and any perceived lasting impact on children's oral language development; and staff practices. The follow-up study was guided by the following RQs:

IPE RQ1: Why have settings continued or discontinued the delivery of Talking Time©?

IPE RQ2: To what extent do settings continue to deliver Talking Time© in accordance with its core components?

IPE RQ3: What have been the facilitators and challenges to sustaining the delivery of Talking Time©?

IPE RQ4: What is the perceived impact of Talking Time© on children's oral language development during this round of delivery?

IPE RQ5: To what extent has Talking Time© led to lasting changes in staff practices, including beyond formal Talking Time© activities?

IPE RQ6: What does the experience of revisited settings indicate about the longer-term sustainability of Talking Time©?

Semi-structured interviews were conducted with six settings between February and April 2026 (three PVIs, two school-based nurseries, and one maintained nursery school), all of which reported continuing Talking Time© activities. In addition, we received email responses from three settings (one PVI and two school-based nurseries). Two of these indicated they had discontinued Talking Time© activities, while the third provided written responses to the interview questions. Interview data were extracted using framework analysis, as described above, with the RQs guiding data extraction. The full follow-up report can be found in Appendix H.

Analysis

Qualitative data

All interviews were recorded (with participant permission) and transcribed verbatim by professional transcribers. Analysis of interview data took place across three team members and utilised a framework analysis approach. Framework analysis is a deductive form of qualitative data analysis with the overarching themes determined before the data is reviewed. It is a useful form of analysis for evaluation research as it facilitates researchers' need to answer specific questions with qualitative data (Klingberg *et al.,* 2023). Framework analysis enabled a systematic approach to organising responses under key IPE dimensions and RQs. Ensuring the RQs posed could be answered was important given they had been developed to align with and test the logic model of the programme. It was also important for rigorously assessing how different contexts (e.g. demographics, available resources, senior leadership support) impact on the assumptions of the logic model. Additionally, this type of analysis is useful for interpreting data efficiently across a team and within a short time frame as it provides clear structure and boundaries for analysis (Goldsmith, 2021). In practice, conducting framework analysis involved placing data from each interview transcript into its assigned row within a matrix and along columns that corresponded with IPE RQs. Once the data input had taken place, data related to specific RQs was analysed more closely to investigate recurring themes and trends across setting types.

Quantitative data

The data extracted from the practitioner's survey were analysed using simple tabulations and frequency distributions. Survey data were triangulated with data collected through interviews in order to validate qualitative findings and highlight where interview data are representative of generalisable or more specific experiences.

Due to small sample sizes, quantitative analysis by setting type was not possible. However, interview data was also able to indicate some potential differences in implementation by setting type.

Table 7: Overview of interview participants

Interview participant	Number	Setting type breakdown	Regional breakdown	Rationale for interview
Practitioners	23	School-based nursery n=13 PVI n=7 Maintained nursery school n=3	East n=7 London n=2 North-West n=3 Yorkshire and the Humber n=6 West Midlands n=5	To 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
Now setting leads	10	School-based nursery n=7 PVI n=2 Maintained nursery school n=1	East n=2 London n=1 North-West n=2 Yorkshire and the Humber n=3 West Midlands n=2	To 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
Control setting leads	7	School-based nursery n=5 PVI n=2	East n=1 London n=2 North-West n=1 Yorkshire and the Humber n=2 West Midlands n=1	To enhance understanding of business as usual in terms of settings' language development policy, pedagogy, and practices
Trainer-mentors	7	All trainer-mentors interviewed	Each trainer mentor covered multipole settings and all mentors were interviewed. No regional breakdown.	To explore their experiences of delivering the training and mentoring sessions and their views on practitioner engagement and any factors influencing that

Table 8: IPE methods overview

Research methods	Data collection methods	Participants / data sources	Data analysis methods	RQs addressed	Implementation / logic model relevance
Qualitative	Semi-structured interviews	Practitioners; setting leads; control setting leads; trainer-mentors	Framework analysis	All	To understand engagement and experience of delivery Talking Time© activities and accessing professional development
Quantitative	Online survey	Setting leads, practitioners	Descriptive statistics	All	To capture practitioner perspective on programme delivery

Costs

In line with the EEF cost evaluation guidance (EEF, 2023), costs are estimated from the perspective of the setting and reported as a per child cost, calculated on the assumption that the intervention is implemented over a three-year period. Costs are disaggregated into pre-requisite, start-up, and recurring costs over this period, and the average annual per child cost of the intervention is then calculated.

Cost data were obtained from the delivery team using excel templates provided by the EEF. Meetings were held between the evaluation and the delivery team to clarify the nature of the data needed. Data were collected using the ingredients approach, in which the costs of all resources necessary to deliver the intervention are identified and valued, regardless of who incurs the cost.

Timeline

Table 9: Timeline

Dates	Activity	Staff responsible / leading
July 2023 – August 2023	Set-up meetings and Intervention Delivery and Evaluation Analysis (IDEA) workshop	Delivery team and evaluation team
September 2023 – November 2023	Ethical approval, develop recruitment materials (including for evaluation preparation study)	Evaluation team and delivery team
January 2024 – July 2024	Recruitment of early years settings (recruitment for both pre-trial and main trial starts in January 2024)	Delivery team with support from evaluation team and SPHs
January 2024 – September 2024	Subsidised delivery period prior to main trial (pilot preparation study)	Delivery team
April 2024 – July 2024	Evaluation activities as part of evaluation preparation study (interviews and test-run of assessments)	Evaluation team
March 2024 – July 2024	Reporting on findings of evaluation preparation study (presentation plus ongoing feedback meetings)	Evaluation team
August 2024 – October 2024	Pre-intervention data collection and baseline assessments	Evaluation team
11 November 2024	Randomisation	Evaluation team
November 2024 – December 2024	Baseline practitioner's survey	Evaluation team
November 2024 – December 2024	Mop-up assessments	Evaluation team
January 2025 – March 2025	IPE planning	Evaluation team
December 2024 – June 2025	Intervention delivery (20 weeks)—includes welcome visits/briefings	Delivery team
March 2025	SAP and IPE design	Evaluation team
April 2025 – June 2025	IPE interviews	Evaluation team
May 2025 – July 2025	Endline assessments	Evaluation team
June 2025 – July 2025	Endline practitioner's survey	Evaluation team
July 2025 – November 2025	Analysis and reporting	Evaluation team
December 2025	First draft of evaluation report	Evaluation team
January 2026	Cost analysis	Evaluation team and delivery team
January 2026 – March 2026	Follow-up to IPE: Returning to ten settings that participated in Talking Time© Conducting one interview per setting, of around 45 minutes in length Evaluation report published on the EEF website	Evaluation team
April 2026	Addendum report on qualitative follow-up study	Evaluation team
June 2026	Final report on main trial, including report on follow-up as an appendix	Evaluation team
July 2026	Publication of evaluation report on the EEF website	The EEF
October 2026	Data submitted to the EEF for archiving	Evaluation team

Impact evaluation results

Participant flow including losses and exclusions

Active recruitment began on 29 January of 2024 and officially ended on 29 July 2024. Recruitment was then extended from July to September 2024 to allow for the replacement of early withdrawals. A total of 2,412 settings were approached to take part in the study. Of these, 644 settings attended information webinars, indicating initial interest in the intervention. Of these 316 settings submitted an expression of interest and 140 subsequently signed an MoU. However, 17 settings withdrew before data collection started in mid-September 2024. Ultimately, 123 settings shared their children's data with the evaluation team and formally joined the study. Among them, 97 were state-funded schools (either maintained nurseries or school-based nurseries) and 26 were PVIs. Settings were randomised with equal probability to the Talking Time© or to the control group, resulting in 62 settings assigned to Talking Time© and 61 to the control group. No settings withdrew after randomisation, and all settings met the minimum requirement of five consented children to participate in the trial.

Across the recruited settings, a total of 2,235 parent/carer consent forms were returned, and 1,923 children were scheduled for the assessments after removing ineligible children and after applying a cap of 15 children per setting. Baseline assessments were conducted between September and November 2024, and approximately 89% (n=1,705) of children scheduled for assessments (n=1,923) were successfully assessed. Among the children not assessed 24% (n=67) were absent at the time of testing. The remaining non-assessments were due to the children being distressed (5%, n=15 children), insufficiently engaged (12%, n=33 children), unable to provide a verbal response (26%, n=74 children), or simply refusing to participate (22%, n=62 children). See the section 'Missing data analysis' for further details.[8]

At follow-up, any child who completed at least one baseline assessment was scheduled for testing.[9] Follow-up assessments were conducted between June and July 2025. A total of 783 children in the Talking Time© group (90.0%) and 757 in the control group (90.9%) were assessed. Reasons for non-assessment at the follow-up differed somewhat from baseline, with 33% (n=55 children) absent on the assessment day (and repeated visits) and the remainder largely attributable to children having left the nursery (43%, n=77 children): refusals (10%, n=18 children); stress (2%, n=4 children); low engagement (3%, n=6 children); and non-verbal response (2%, n=4 children) were less common than at baseline.

Since children did not necessarily complete all the assessments at either baseline or follow-up, the total number of usable observations varies across assessment types. Moreover, the primary outcome is a composite index requiring children to have completed three specific assessments at both time points, further reducing the analytical sample. In total, 1,439 children had complete data for the primary outcome, comprising 725 in the Talking Time© group and 714 in the control group.

[8] Note that the number of missing cases varies with the assessment, as some children took some assessments but not others. In this paragraph and in the following we used the number of missing cases reported for the REV assessment.
[9] This group includes children who did not complete the GAPS test because of the stopping rule and that were scored 0.

Figure 2: Participant flow diagram

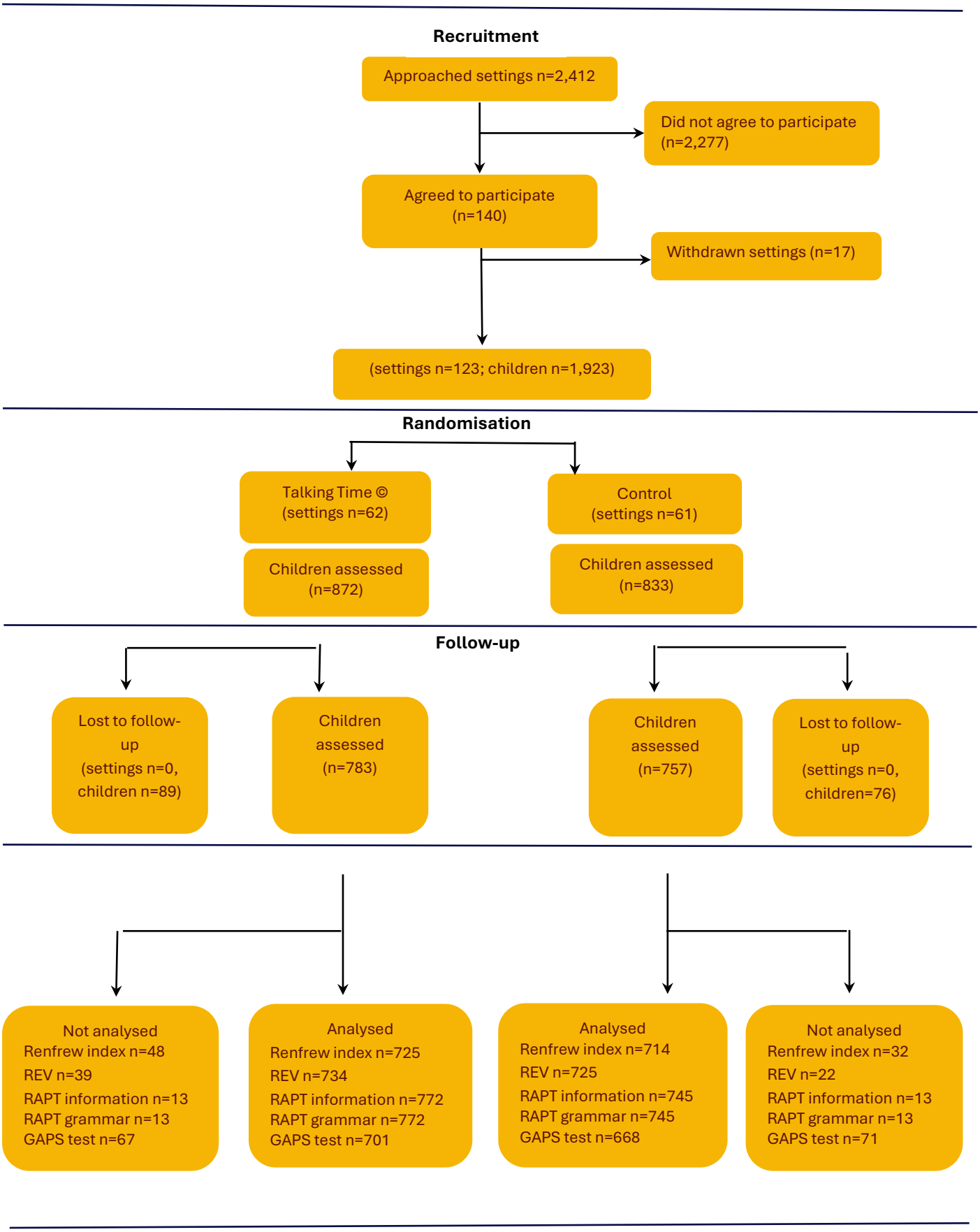


Table 10: MDES at different stages

		Protocol		Randomisation		Analysis	
		Overall	EYPP	Overall	EYPP	Overall	EYPP
MDES		0.17	0.29	0.20	0.36	0.17	0.36
Pre-test/post-test correlations	Level 1 (child)	0.40	0.40	0.40	0.40	0.82	0.82
ICCs	Level 2 (class)	0.10	0.10	0.10	0.10	0.11	0.11
Alpha		0.05	0.05	0.05	0.05	0.05	0.05
Power		0.8	0.8	0.8	0.8	0.8	0.8
One-sided or two-sided?		Two-sided	Two-sided	Two-sided	Two-sided	Two-sided	Two-sided
Average cluster size		15	3	13	2	12	2
No. of settings	Intervention	65	65	62	62	62	62
	Control	65	65	61	61	61	61
	Total:	130	130	123	123	123	123
No. of children	Intervention	975	195	977	168	725	135
	Control	975	195	946	175	714	121
	Total:	1,950	390	1,923	343	1,439	256

Table 10 reports the MDES for the full sample and for the EYPP subgroup, alongside the parameters observed in the data that were used for the power calculations. At the analysis stage, the MDES is 0.17 SD for the full sample and 0.36 SD for the EYPP subgroup. Thus, while the trial is adequately powered to detect impacts for the overall cohort, it is not powered to detect plausible effects for EYPP children.

Table 10 also shows how the expected MDES evolved over time, from the protocol stage through randomisation, as sample size and parameter assumptions changed. Although the final sample for the analysis was smaller than initially anticipated, the pre-test/post-test correlation was higher than assumed at the design stage. These factors offset each other, resulting in an MDES at the analysis stage which is broadly similar to the estimate predicted at the protocol stage.

Attrition

Table 11: Child-level attrition from the trial (primary outcome)

		Intervention	Control	Total
No. of children	Randomised	872	833	1,705
	Analysed	725	714	1,439
Child attrition (from randomisation to analysis)	Number	147	119	266
	Percentage	16.7%	14.3%	15.6%

Overall attrition, defined as the proportion of children randomised who were not included in the primary outcome analysis using the composite Renfrew index, was 15.6% (266 out of 1,705 children assessed at baseline) with a slightly larger percentage (16.7%) in the intervention group (Table 11). This number includes children that were randomised and not assessed at baseline. The percentage of children who were assessed at baseline and that were not assessed at follow-up

was 10.8% (174 out of 1,439 children). This overall figure reflects approximately 10% non-response rate at follow-up for each specific test. The percentage of children assessed by any of the tests at baseline that were not assessed by any of the tests at follow-up was 9.7% (165 of 1,705 children). The attrition rate for the primary outcome (the Renfrew index) becomes larger than 10% because the primary outcome is a composite which requires completion of two distinct assessments to be calculated. At follow-up absence accounted for 33% of non-completion (n=55 children), but the most common reason became children having left the nursery (43%, n=77 children), while distress (2%, n=4 children), lack of engagement (3%, n=6 children), and refusal (10%, n=18 children) were less prominent than at baseline.

Children and setting characteristics

Table 12 presents the means of setting-level characteristics, child-level characteristics, and child assessment scores at baseline. The means are calculated over all children for whom at least one assessment was available at baseline.

Setting-level and child-level characteristics are very similar across the intervention and the control groups, indicating that randomisation was successful. Baseline assessment scores are slightly higher in the intervention group than in the control group. However, for the primary outcome and for all other assessments outcomes (with the exception of the RAPT assessment) the standardised mean differences are smaller than 0.1 SD.

The proportion of children eligible for EYPP in our study sample is 19%. Nurseries do not routinely report EYPP status, and we collected the information from practitioners. EYPP status was recorded at baseline in September 2024 and again at follow-up in June 2025 at endline, because EYPP eligibility was often not known at the beginning of the academic year. The proportion of EYPP children observed in the sample is slightly higher than the national average. According to the DfE, around one in seven children in nurseries were in receipt of the EYPP in 2025, corresponding to 14% of the population.[10] The delivery team made considerable efforts during recruitment to include nurseries with a higher likelihood of enrolling EYPP children. At the randomisation stage, we set-up and implemented an algorithm to prioritise the inclusion of EYPP-eligible children whenever the number of eligible children in a setting exceeded the maximum target number of assessments per setting (15).

The proportion of children with EAL in our study sample is 34%, which is slightly larger than the national average. The DfE reported that in the academic year 2024/2025, 30% of children in nursery schools were recorded as having a first language known or believed to be different from English.

Appendix I reports the same table for the sample used in the analysis. As attrition between baseline and follow-up exceeds 10%, we compare the distribution of the baseline covariate in the analytical sample, comprising children for whom we observe the primary outcome at both baseline and follow-up. The results are very similar to those observed in the full baseline sample. Demographic characteristics are nearly identical between the treatment and control groups, while baseline assessment scores are slightly larger in the intervention group. Overall, there is no indication that missing data have introduced substantial bias into the analytical sample. This issue is discussed in more detail in the section on 'Missing data analysis'.

Study settings are disproportionately located in more economically deprived areas. Figure 3 shows the distribution of the IDACI index for Talking Time© settings (bars), compared to the distribution across all UK Lower Layer Super Output Areas (LSOAs; small geographic subdivision units of which there are about 36,000 in the UK). The average IDACI score for Talking Time© settings is 0.23 (median 0.22), compared with an average of 0.16 (median 0.13) in the overall population.

[10] See: https://explore-education-statistics.service.gov.uk/find-statistics/funded-early-education-and-childcare/2025?utm_source=chatgpt.com

Table 12: Baseline characteristics of groups as randomised

Setting level (categorical)	National-level mean	Intervention group		Control group		
		n/N (missing)	Count (%)	n/N (missing)	Count (%)	
PVIs	68.4%[a]	62/62 (0)	19%	61/61 (0)	19%	
Setting level (continuous)		n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)	
IDACI score	0.16[b]	62/62 (0)	0.23 (0.11)	61/61 (0)	0.23 (0.11)	
Child level (categorical)		n/N (missing)	Count (%)	n/N (missing)	Count (%)	
EYPP-eligible	14%[c]	165/852 (20)	19%	143/809 (24)	19%	
EAL	30.2%[d]	277/864 (8)	34%	272/825 (8)	35%	
Child level (categorical)		n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)	
Female		435/872 (0)	0.49 (0.50)	420/833(0)	0.49 (0.50)	
Child level (continuous)		n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)	Effect size (SE)
Age in months		872 (0)	42.9 (4.0)	833 (0)	42.9 (4.1)	
Renfrew index[e] (primary outcome)		817/872 (55)	13.6 (6.8)	796/833 (37)	13.2(6.8)	0.06 (0.05)
REV[e]		829/872 (43)	17.8 (9.5)	809/833 (24)	17.6 (9.6)	0.02 (0.05)
RAPT information[e] (vocabulary knowledge)		858/872 (14)	14.0 (6.9)	820/833 (13)	13.3 (6.8)	0.11 (0.05)
RAPT grammar[e]		858/872 (14)	8.6 (6.2)	820/833 (13)	8.3 (6.1)	0.05 (0.05)
GAPS test[e]		796/872 (76)	3.5 (4.0)	749/833 (84)	3.4 (3.9)	0.02 (0.05)

[a] *Based on the 2025 Childcare and Early Years provider survey. In 2025, there were approximately 21,400 group-based providers, and 9,900 school-based providers. See: https://explore-education-statistics.service.gov.uk/find-statistics/childcare-and-early-years-provider-survey/2025?utm_source=chatgpt.com.*

[b] *Averages of LSOAs in England.*

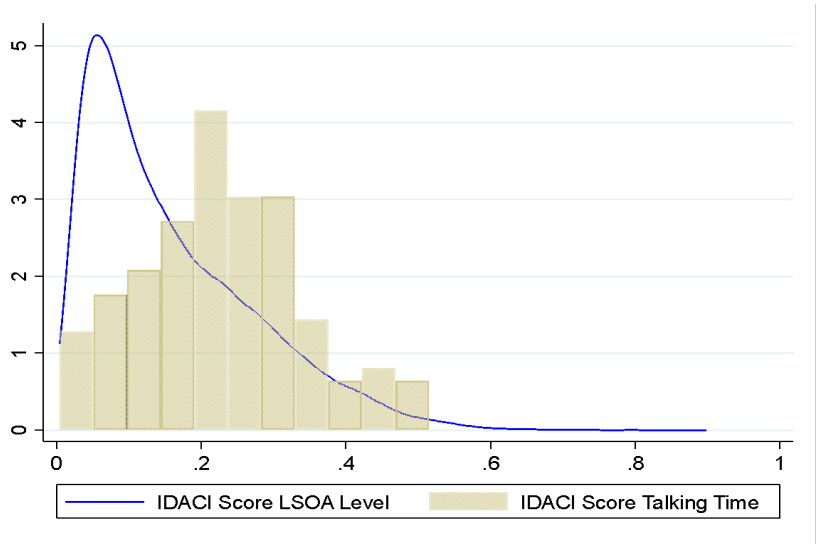
[c] *In 2024, based on DfE statistics. See: https://explore-education-statistics.service.gov.uk/find-statistics/funded-early-education-and-childcare/2024?utm_source=chatgpt.com.*

[d] *Recorded as having EAL in nursery schools according to DfE statistics. See: https://dera.ioe.ac.uk/id/eprint/41289/1/Schools%2C%20pupils%20and%20their%20characteristics%2C%20Academic%20year%20202425%20-%20Explore%20education%20statistics%20-%20GOV.UK.pdf?utm_source=chatgpt.com.*

[e] *There are no population-level estimates of average scores of Renfrew and GAPS test assessments in England.*

SE=standard error.

Figure 3: Setting-level IDACI scores



An online baseline survey was conducted to collect setting-level information from nursery leaders. Respondents were the main contact at the time of recruitment, and were most commonly headteachers, managers, or class teachers. After removing duplicate responses from the same school, we obtained a total of 99 responses out the 123 settings (84%), of which 46 were in the intervention group and 53 in the control group. About 80% of respondents (72) categorised their settings as primary schools with nurseries, 14% as day nurseries (14 respondents), 9% as preschools (nine respondents), and 7% as nursery schools (eight respondents). The sample included 48 state-maintained nurseries (47%), 32 academies (31%), 15 non-profit nurseries (13%), eight private nurseries (8%), and two independent nurseries (2%). In most cases (84%, n=86 respondents), settings operated as single sites, and the remaining 16% (n=16 respondents) were part of multisite nursery chains or groups.

Table 13 report key statistics observed in the survey including a comparison between the treatment and control group. On average, settings reported about seven staff involved in the delivery of early education and care provision. The average child-to-staff ratio was 8.9, and only 67% of settings reported having sufficient staff to meet current demand. These figures suggest an average nursery size of approximately 65 places per nursery, which is aligned with averages reported in national statistics. Of the staff members, 14% were employed part-time, with the remainder permanently employed. In terms of qualifications, 60% of staff (as reported by the leaders responding to the survey) held Level 3–5 qualifications, while 28% held qualification Level 6 or above.

Talking Time© was not the only language intervention implemented in these settings. At baseline, 57% of settings reported plans to deliver additional language development interventions in the forthcoming academic year. Interventions cited included: Colourful Semantics; Concept Cat; Every Child a Talker (ECAT); Early Talk Boost; Language Enrichment Activity Programme (LEEP;) Tales Toolkit; and WellComm.

Across all indicators considered, the differences between intervention and control groups are minimal. Since not all settings responded to the survey (particularly in the intervention group) any remaining differences may reflect missing data rather than true underlying differences.

Table 13: Key setting characteristics

	Intervention group (SD)	Control group (SD)	All settings (SD)
Average number of staff currently employed	8.00 (5.6) N=46	7.19 (4.56) N=53	7.6017.5% (5.07) N=103

Proportion of part-time staff currently employed	0.18 (0.25) N=46	0.11 (0.15) N=53	0.14 (0.20) N=103
Proportion of staff with Level 3–5 qualifications	0.58 (0.30) N=46	0.59 (0.24) N=53	0.60 (0.27) N=103
Proportion of staff with Level 6 or above	0.26 (0.16) N=46	0.29 (0.18) N=53	0.28 (0.17) N=103
Average number of children per staff member	9.24 (4.17) N=45	8.77 (3.05) N=53	8.92 (3.55) N=102
Proportion believing having sufficient staff	0.70 (0.47) N=46	0.66 (0.48) N=53	0.67 (0.47) N=103
Proportion with language intervention other than Talking Time©	0.52 (0.51) N=46	0.62 (0.49) N=53	0.57 (0.50) N=103

Outcomes and analysis

Estimation of effect sizes

Primary analysis

Figure 4 shows the histograms of the primary outcome (the composite Renfrew index) at baseline and at follow-up for all children in the analytical sample. The charts show the average improvement in scores over time as the distribution shifts to the right. The distributions also appear more clustered at follow-up indicating a reduction in the variance. The distributions are approximately bell-shaped, although statistical test reject the hypothesis of normal. The primary analysis uses a multilevel model, which handles cluster effects by design, so that no further correction for clustering of observations is needed. After estimating the multilevel models with random intercepts, we calculated conditional level-1 residuals and level-2 (setting level) random effect residuals using empirical Bayes estimates. We then inspected the distribution of the residuals using histograms and Quantile-Quantile (Q-Q) plots. We did not find substantial departures from normality.

Figure 4: Histogram of test scores at baseline and follow-up

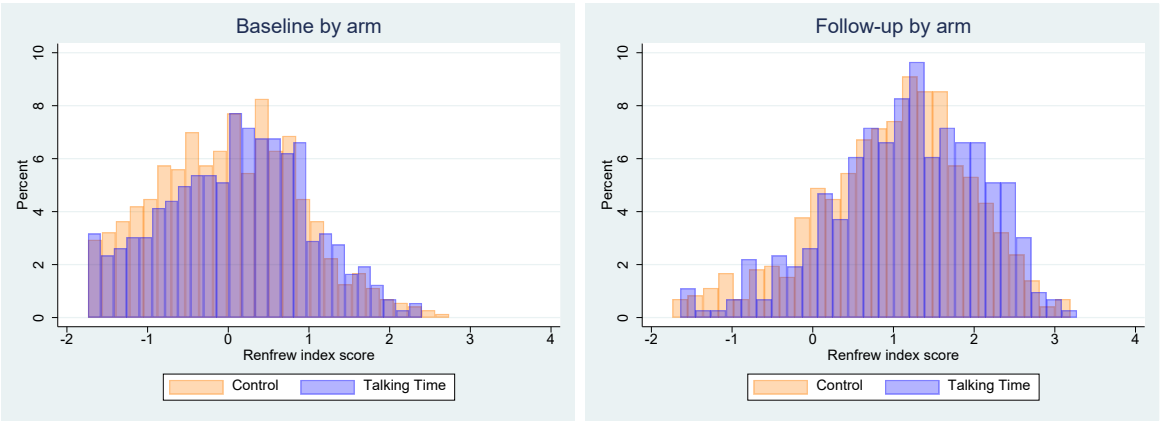


Table 14 reports the unadjusted means for the primary outcome and the standardised impact of the intervention on the primary outcome. The latter is obtained from the estimation of a multilevel model including the pre-test scores and the stratification dummies as covariates. The variances used to calculate the standardised Hedges' g results are reported in Table 15. Talking Time© had a positive impact on the Renfrew composite index of 0.09 SD corresponding to one month's

progress, with the 95% confidence interval (CI) entirely above zero. The index capture improvements across different dimensions of language development, consistent with the logic model that predicts distinct intervention activities will have positive effects on different language abilities.

Table 14: Impact of Talking Time© on the primary outcome

Outcome	Unadjusted means				Effect size		
	Intervention group		Control group				
	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention; control)	Hedges' g (95% CI)	P-value
Renfrew composite index	725 (0)	1.15 (1.09,1.22)	714 (0)	0.99 (0.92,1.06)	1,439	0.083 (0.001, 0.164)	0.047

Table 15: Effect size estimation (primary outcome)

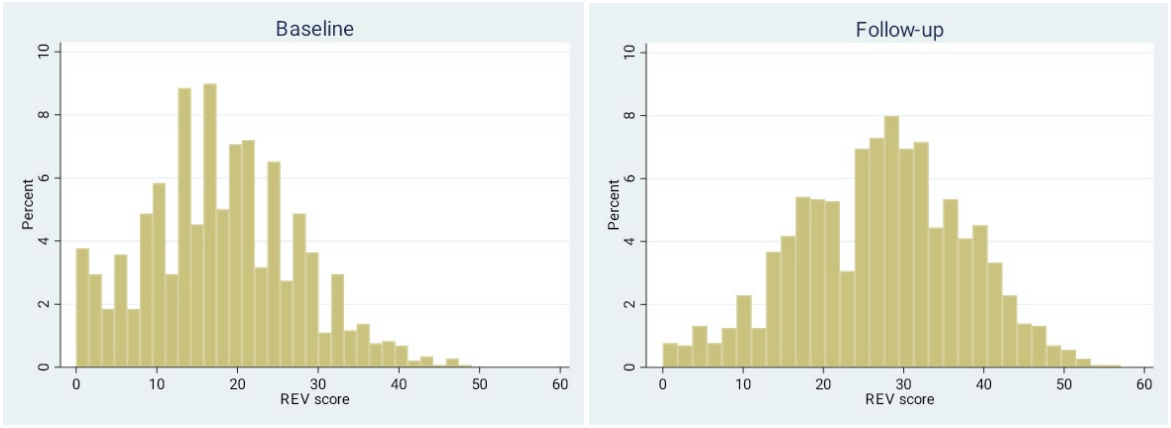
Outcome	Unadjusted differences in means	Intervention group		Control group		Variance of outcome	Pooled variance
		Adjusted differences in means	n (missing)	Variance of outcome	n (missing)		
Renfrew composite index	0.165	0.083	725 (0)	0.792	714 (0)	0.800	0.800

Secondary analysis

Figure 5 shows the histograms of all assessments at baseline and at follow-up for the children in the analytical sample. REV and RAPT are bell-shaped, although they fail a normality test. There is some evidence of floor effects at the baseline for the RAPT (information) assessment and for the GAPS test assessment.

This is most visible in the case of the GAPS test, for which there were 40% of zero observations, which is largely explained by the assignment of zero scores to the stopping rule adopted during assessment administration at baseline. The floor effects disappear at follow-up including for the GAPS test.

Figure 5: Histogram of test scores at baseline and follow-up



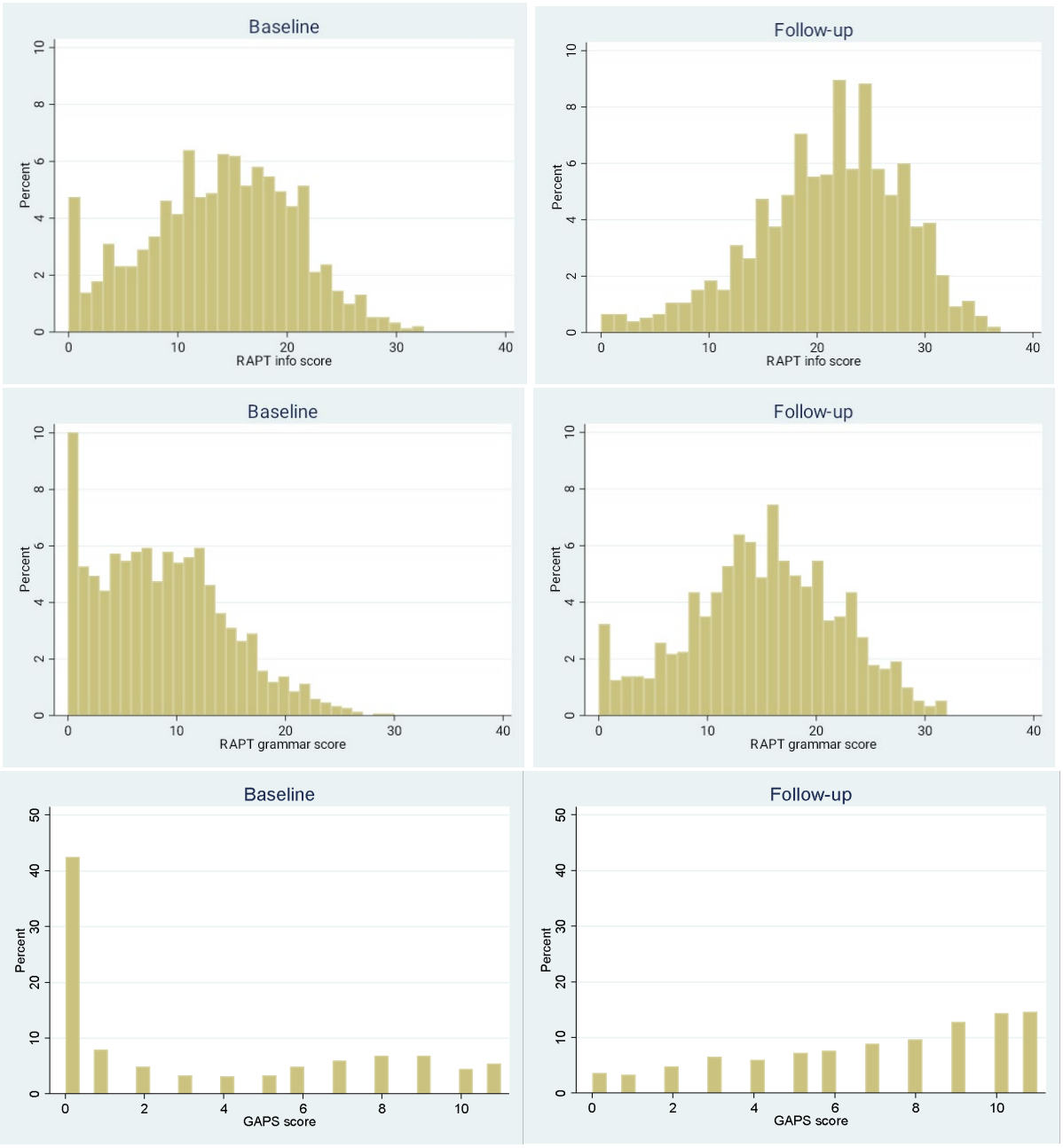


Table 16 presents the estimated impact of Talking Time© on the four secondary assessments, while Table 17 reports the variances that were used in the calculation of the standardised effect sizes. Talking Time© had a positive impact on all three Renfrew assessments, although only the CI for the expressive vocabulary assessments lies entirely above zero. By contrast, the impact of Talking Time© on the GAPS test assessment was negative. Overall, these findings suggest that the intervention had a stronger impact on vocabulary development than on grammar abilities.

Table 16: Impact of Talking Time© on the secondary outcomes

Outcome	Unadjusted means				Effect size		
	Intervention group		Control group				
	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (Intervention; control)	Hedges' g (95% CI)	P-value
REV	734 (95)	27.61	725 (84)	26.38	1,459	0.079 (0.002, 0.157)	0.044

RAPT information (vocabulary knowledge)	772 (86)	21.60	745 (75)	20.53	1,517	0.064 (-0.014, 0.142)	0.248
RAPT grammar	772 (86)	15.57	745 (75)	14.74	1,517	0.092 (-0.20, 0.203)	0.107
GAPS test	701 (95)	6.93	668 (81)	7.09	1,369	-0.054 (-0.136, 0,028)	0.194

Table 17: Effect size estimation (secondary outcome)

Outcome	Unadjusted differences in means	Intervention group		Control group		Variance of outcome	Pooled variance
		Adjusted differences in means	n (missing)	Variance of outcome	n (missing)		
Expressive vocabulary	1.230	0.760	734 (95)	107.72	725 (84)	112.35	110.32
RAPT information	1.068	0.439	772 (86)	47.13	745 (75)	47.14	47.39
RAPT grammar	0.827	0.565	772 (86)	45.93	745 (75)	44.88	45.56
GAPS test	-0.167	-0.214	701 (95)	10.33	668 (81)	10.55	10.43

In principle, this result could be driven by the floor effects in the GAPS test scores observed at baseline. To assess this possibility, we re-estimated the impact of Talking Time© on GAPS test outcomes, using a covariate in the ANCOVA model, the baseline Renfrew index in place of baseline GAPS test scores, and additionally estimated the model using multiple imputation. In both cases, the results were very similar to those obtained with the original model specification, indicating that the findings are unlikely to be driven by baseline floor effect.

Subgroup analysis

Table 18 reports the impact of Talking Time© on the primary outcome for the subgroups of interest: EYPP children; and children with EAL. Table 18 also reports the unadjusted means, sample sizes for the subgroups, and Hedges' g. The impact of Talking Time@ was slightly larger among EYPP children and considerably larger among children with EAL. The samples used in the estimation, however, are very small. In Appendix J, we report the results of a subgroup analysis using interaction models that allows testing the difference in the effect size among the subgroups. This analysis does not find significant differences in effect sizes of children eligible and non-eligible for EYPP, and between children with and without EAL.

Table 18: Impact of Talking Time© on the primary outcome for EYPP children and children with EAL

Outcome	Unadjusted means				Effect size		
	Intervention group		Control group				
	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention; control)	Hedges' g (95% CI)	P-value
Renfrew index (EYPP)	135 (6)	1.000 (0.954, 1.246)	121 (17)	0.867 (0.696, 1.038)	256 (135; 121)	0.103 (-0.043,0.249)	0.168
Renfrew index (EAL)	219 (0)	0.590 (0.458, 0.722)	220 (7)	0.414 (0.290, 0.539)	439 (219; 220)	0.151 (0.026,0.277)	0.018

Table 19 and Table 20 report the subgroup effects for all secondary outcomes for EYPP children and for children with EAL, respectively. The results for the EYPP subgroup are mixed, with two positive effects and two negative effects. All CIs are wide and contain zero. Please note again that the sample size of the study was such as to be able to detect only very large effects on subgroups of the population as discussed in the section "Participant flow including losses and exclusions".

Table 19: Subgroup analysis EYPP children

Outcome	Unadjusted means				Total n (intervention; control)	Effect size	
	Intervention group		Control group				
	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)		Hedges' g (95% CI)	P-value
REV (EYPP)	136	26.60 (25.10, 28.09)	123	24.59 (22.77, 26.42)	259 (136, 123)	0.084 (-0.144,0.312)	0.334
RAPT information (vocabulary knowledge) (EYPP)	143	21.32 (20.15, 22,48)	127	20.16 (18.84, 21.48)	260 (143, 127)	-0.064 (-0.290,0.162)	0.509
RAPT grammar (EYPP)	143	15.04 (13.90, 16.18)	127	13,66 (12,49, 14.82)	260 (143, 127)	0.193 (-0.034,0.419)	0.080
GAPS test (EYPP)	126	6.18 (5.66, 6.71)	108	6.87 (6.25, 7.49)	234 (126, 108)	-0.090 (-0.326,0.146)	0.377

The impact of Talking Time© on children with EAL is positive for all assessments. CIs, however, are large and include zero with only one exception (RAPT). Notably, while the largest effects in the full sample were observed for expressive vocabulary (REV), for EYPP children and those with EAL, the largest effects are found in grammar (RAPT), suggesting that these subgroups may benefit from the intervention in distinct ways.

Table 20: Subgroup analysis children with EAL

Outcome	Unadjusted means				Total n (intervention ; control)	Effect size	
	Intervention group		Control group				
	N (missing)	Mean (95% CI)	N (missing)	Mean (95% CI)		Hedges' g (95% CI)	P-value
REV (EAL)	222	21.50 (20.18, 22.83)	225	20.19 (18.93, 21.45)	447 (222, 225)	0.110 (-0.061,0.282)	0.064
RAPT information (vocabulary knowledge) (EAL)	236	17.75 (16.82, 18.67)	230	16.84 (15.93, 17.74)	466 (236, 230)	0.143 (-0.026,0.312)	0.089
RAPT grammar (EAL)	236	12,39 (11,49, 13,28)	230	11.43 (10.59, 12.28)	466 (236, 230)	0.177 (0.007,0.346)	0.059
GAPS test (EAL)	219	5.89 (5.44, 6.08)	214	5.64 (5.20, 6.09)	433 (219, 214)	-0.021 (-0.196,0.153)	0.782

To illustrate the uncertainty surrounding the results, we apply a procedure to adjust for multiple hypotheses testing using a FDR correction. When many coefficients are estimated at the same time, some may obtain a low probability value purely by chance. Table 21 reports the coefficient estimates from the subgroup analysis, their associated p-values, and the corresponding FDR-adjusted significance thresholds. The tests are ordered by increasing order of their p-values. None of

the p-values fall below the corresponding FDR-adjusted threshold, suggesting that no effect is statistically 'relevant' once we account for multiple testing.

Table 21: FDR-adjustment for multiple testing of hypotheses

	Effect size	P-value	FDR threshold
Renfrew composite index (EAL)	0.151	0.018	0.010
RAPT grammar (EAL)	0.177	0.059	0.020
REV (EAL)	0.110	0.064	0.030
RAPT grammar (EYPP)	0.193	0.080	0.040
RAPT information (vocabulary knowledge) (EAL)	0.143	0.089	0.050
Renfrew composite index (EYPP)	0.103	0.168	0.060
REV (EYPP)	0.084	0.334	0.070
GAPS test (EYPP)	-0.090	0.377	0.080
RAPT Information (vocabulary knowledge) (EYPP)	-0.064	0.509	0.090
GAPS test (EAL)	-0.021	0.782	0.100

Analysis in the presence of non-compliance

Talking Time© consists of 15-minute sessions delivered twice a week for 20 weeks and embedded within normal nursery practice. One of the 20 weeks included three sessions, rather than two sessions, and the maximum number of sessions attended by each child is therefore 41. Attendance at Talking Time© sessions was high. Figure 6 shows the number of sessions attended by children who were assessed at follow-up. Most children attended at least 30 sessions, with a relatively small number attending fewer sessions.

Figure 6: Percentage of Talking Time© sessions attended

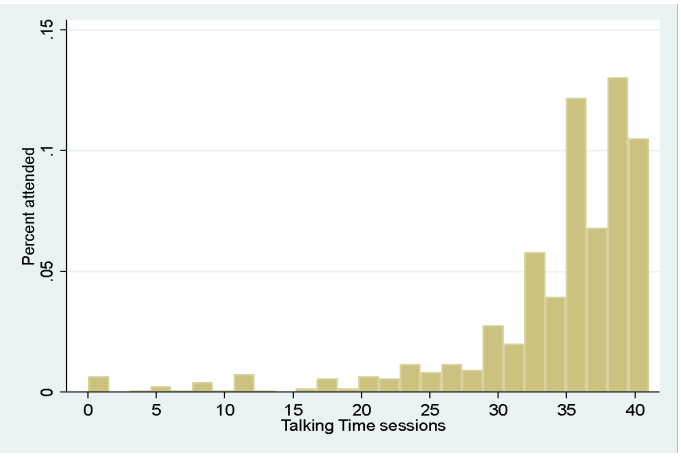


Table 22 reports the average proportion of sessions attended. In the analysis sample, children attended on average 83% of sessions. The percentage of sessions attended is smaller in the sample of children scheduled for the assessment at baseline (target sample) and for the sample of children that were assessed at baseline (baseline sample). Children who were not assessed at baseline were less likely to attend the sessions, and some children left the nursery between baseline and follow-up assessment. Attendance rates were very similar across session types, although attendance was slightly lower for Word Play and Hexagons sessions.

Table 22: Proportion of Talking Time© sessions attended

Talking Time© sessions	Target sample n=977 (SD)	Baseline sample n=872 (SD)	Analysis sample n=783 (SD)
All sessions	0.71 (0.33)	0.79 (0.24)	0.83 (0.18)
Story Conversations	0.72 (0.33)	0.81 (0.23)	0.84 (0.17)
Word Play	0.71 (0.36)	0.80 (0.28)	0.8 (0.23)
Hexagons	0.68 (0.36)	0.77 (0.29)	0.80 (0.24)

Figure 7 shows that attendance did not vary significantly over the academic year. Attendance was marginally lower during the first session and the last session, which may reflect general nursery attendance patterns. Overall, attendance remained stable throughout the year.

Figure 7: Attendance of Talking Time© sessions over the academic year

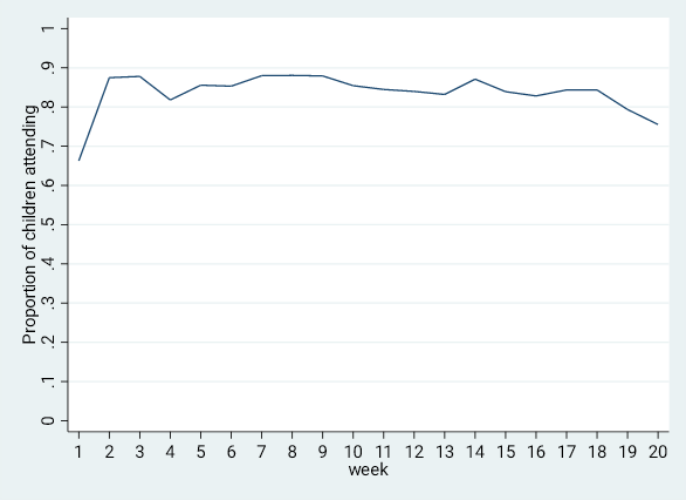


Table 23 presents for each assessment the CACE effects estimated for each assessment using the procedure described in the 'Methods' section. The coefficient represents the impact of attending one additional Talking Time© session on standardised test scores. The impact is local to children who attended the sessions and cannot be extrapolated to the overall population. Multiplying the per session impact by the average attendance yields an effect very similar to the estimated impact observed for the full population. The impact on compliers is not substantially larger than the average population impact, reflecting the fact that attendance was high and that most children attended the majority of the sessions. Consequently, the compliers do not represent a highly selected subgroup for whom the intervention had disproportionately large effects.

Table 23: CACE effects of one additional session

	Sample	Coefficient	Standard error	P-value
Renfrew index	1,416	0.002	0.001	0.057
Expressive vocabulary	1,435	0.002	0.001	0.077
RAPT information (vocabulary knowledge)	1,493	0.002	0.002	0.294
RAPT grammar	1,493	0.003	0.002	0.104
GAPS test	1.348	-0.001	0.001	0.262

Next, we look at the impact of the intervention at varying levels of attendance. As mentioned above, there are no theoretical grounds for defining a specific attendance cut-off. We therefore, estimate CACE effects across all possible cut-offs. The proportion of children associated with each potential cut-off is shown in Figure 8. The share of children attending at least a given number of sessions declines sharply after 30 to 35 sessions and remains relatively stable below this threshold.

Figure 8: Proportion of children attending a given number of sessions

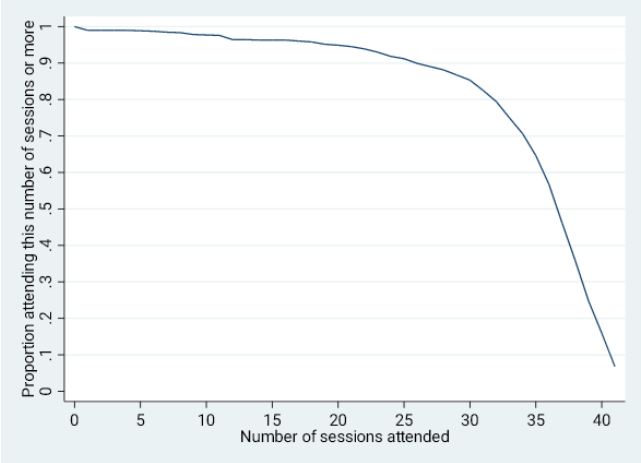
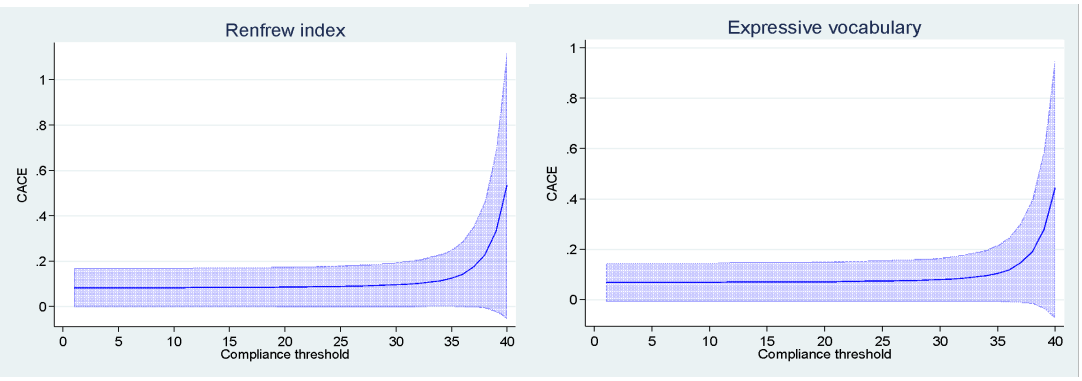
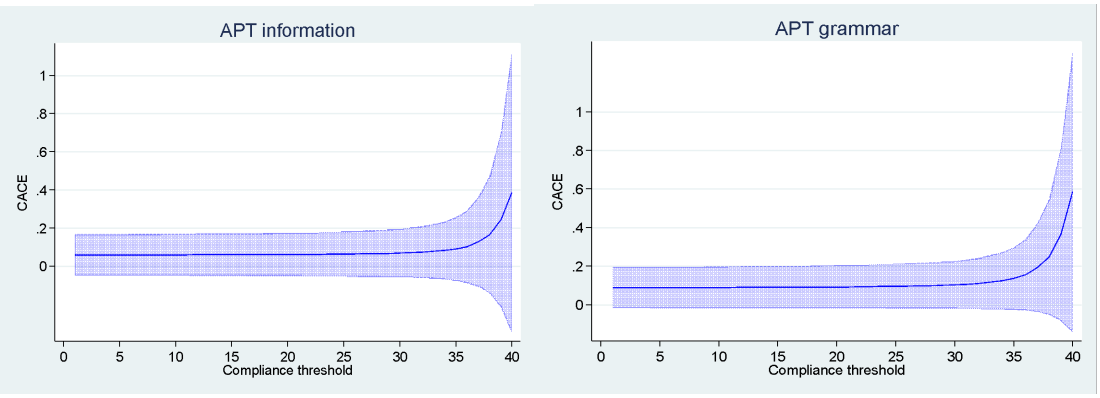
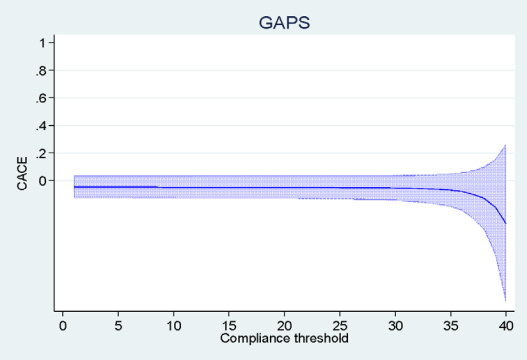


Figure 9 plots the CACE effects for varying attendance thresholds for each assessment. These estimates should not be interpreted as dose–response relationships. The estimates are local to the population attending at least a given number of sessions, and the characteristics of this population change as threshold increases. For example, children attending nearly all sessions may also have more continuous nursery attendance and may differ in unobserved ways that influence the response to the intervention. With these caveats in mind, the estimated effects are more stable and similar for the large majority of the sample of children attending fewer than 35 sessions. Effects appear larger for children attending more than 35 sessions, however, these estimates are more uncertain and apply to a highly selected subgroup whose response may be driven in part by unobserved characteristics.

Figure 9: CACE effects for all levels of attendance







Missing data analysis

We begin the missing data analysis by examining the proportion of missing outcome observations at both baseline and follow-up. Baseline missingness is defined as the proportion of children who were scheduled for the assessments but were not assessed. Follow-up missingness is defined as the proportion of children who were not assessed at follow-up among those who were assessed at baseline (the randomisation sample). The mechanisms underlying missingness in the two cases differ and have distinct implications for the study. Missingness at baseline primarily relates to the representativeness of the sample, whereas missingness at follow-up is more directly associated with the risk of attrition bias.

The proportion of missing cases for all assessments (including the composite index) are reported in Table 24. The proportion of missing cases at baseline was relatively high particularly for the GAPS test. At follow-up the proportion of missing cases is lower and does not differ substantially across assessments.

Table 24: Missing outcome cases at baseline and follow-up (%)

	Missing at baseline n=1,923 (missing cases)	Missing at follow-up n=1,613, n=1,638; n=1.678; n=1,545[a] (missing cases)
Renfrew index	16.1% (310)	10.8% (174)
REV	14.8% (285)	10.9% (179)
RAPT	12.7% (245)	9.6% (161)
GAPS test	19.7% (378)	11.4% (176)

[a] *The numbers refer to the samples of children that were assessed at baseline. The samples vary for each test because both at baseline and follow-up children took some assessments but not others.*

Table 25 and Table 26 report the reasons for missing outcome observations reported by the assessors at baseline and at follow-up, respectively. At baseline, across the three tests, child absence accounted for approximately 25% of missing cases. Around half of missing observations were due to children unable to provide a verbal response or refusing to participate. Distress and lack of engagement accounted for a further 10–15% of missing cases. These patterns highlight the

challenges of conducting assessments in early years settings. Increasing the number or length of assessment visits would likely not have substantially reduced missingness.

Table 25: Reported reasons for missing outcomes at baseline

	REV n=285 (no. of cases)	RAPT n=245 (no. of cases)	GAPS test n=378 (no. of cases)
Child absent or unwell	23.5% (67)	26.9% (66)	17.5% (66)
Child had to leave	3.2% (9)	1.2% (3)	1.6% (6)
Child was distressed	5.3% (15)	4.5% (11)	5.0% (19)
Child was disengaged	11.6% (33)	5.3% (13)	11.9% (45)
No verbal response	26.0% (74)	40.4% (99)	36.5% (138)
Child refused	21.8% (62)	14.3% (35)	20.1% (76)
Other	8.8% (25)	7.4% (18)	7.4% (28)

The reasons for missing observations differ markedly at follow-up. Refusals, lack of verbal response, disengagement, and distress together account for less than 20% of missing cases. By this point, children are nearly one year older and more willing to complete the assessment. In addition, only children who had been assessed at baseline were eligible for assessment at follow-up, meaning that some of the more 'difficult' cases (e.g. children unable to provide a verbal response) had already exited the sample. Child absence accounts for about 35% of missing cases at follow-up, possibly reflecting the proximity to the holiday period and the shorter assessment window compared with the baseline. The most common reason for missing data at follow-up is the child having left the nursery, which accounts for over 40% of missing cases (77 cases for REV, 74 for RAPT, and 75 for GAPS).

Table 26: Reported reasons for missing outcomes at follow-up

	REV n=179	RAPT n=161	GAPS test n=176
Child absent or unwell	33.0 (59)	38.5% (62)	33.5% (59)
Child had to leave	2.8% (5)	0.6% (1)	0.6 (1)
Child was distressed	2.2% (4)	1.2% (2)	3.4% (6)
Child was disengaged	3.4% (6)	1.9% (3)	1.7% (3)
No verbal response	2.2% (4)	4.4% (7)	4.0% (7)
Child refused	10.1% (18)	3.1% (5)	10.2% (18)
Child left the nursery	43.0%	45.9%	42.6%

	(77)	(74)	(75)
Other	3.4 (6)	4.4% (7)	4.0% (7)

Table 27 and Table 28 compare the proportion of missing cases between the intervention and the control group at baseline and follow-up. The proportions are very similar across groups at both time points. Although a slightly higher proportion of children completed the assessments in the intervention group, t-tests indicate that these differences are not statistically significant. We therefore find no evidence of differential attrition between the two groups. Missingness appears to be independent of treatment assignment and of programme implementation.

Table 27: Proportion of missing cases in the intervention and control group (baseline)

	REV N=1,923 (missing cases)	RAPT N=1,923 (missing cases)	GAPS test N=1,923 (missing cases)
Intervention group	0.151 (148)	0.122 (119)	0.185 (181)
Control group	0.145 (137)	0.133 (126)	0.208 (197)
Difference	0.007	-0.011	-0.023
T-statistic	0.411	0.749	1.268
P-value	0.681	0.454	0.205

Table 28: Proportion of missing cases in the intervention and control group (follow-up)

	REV N=1,638 (missing cases)	RAPT N=1,678 (missing cases)	GAPS test N=1,545 (missing cases)
Intervention group	0.115 (95)	0.100 (86)	0.119 (95)
Control group	0.104 (84)	0.091 (75)	0.108 (81)
Difference	0.011	0.009	0.011
T-statistic	0.698	0.609	0.692
P-value	0.485	0.542	0.489

Table 29 and Table 30 present the results of regressions of missingness on observed child characteristics. Separate regressions are estimated for the intervention and control groups, for each of the three tests, and for baseline and follow-up data. Coefficients are reported as odds ratio and represent the likelihood that an observation is missing. At baseline, female children and older children are significantly less likely to have missing outcomes. Each additional month of age substantially reduces the probability of missing any assessment. Children eligible for EYPP and children with EAL are approximately twice as likely to have missing test data compared to other children. These effects are highly statistically significant and consistent across assessments.

Table 29: Determinants of missingness in the intervention and control groups (baseline)

	REV		RAPT		GAPS test	
	Control group N=914 Coeff. (SE)	Intervention group N=951 Coeff. (SE)	Control group N=914 Coeff. (SE)	Intervention group N=951 Coeff. (SE)	Control group N=914 Coeff. (SE)	Intervention group N=951 Coeff. (SE)
Gender	0.65** (0.13)	0.54** (0.10)	0.65** (0.14)	0.64** (0.13)	0.49*** (0.09)	0.61** (0.11)
Age	0.97 (0.02)	0.96* (0.02)	0.94** (0.02)	0.94** (0.02)	0.94** (0.02)	0.94** (0.02)
EYPP	1.89** (0.46)	0.96 (0.25)	1.93** (0.49)	0.95 (0.27)	1.99*** (0.42)	1.08 (0.25)
EAL	1.83** (0.39)	1.82** (0.37)	2.11*** (0.46)	2.17*** (0.48)	1.39* (0.26)	1.56** (0.29)
IDACI score	1.37 (1.84)	1.53 (1.59)	5.11 (6.91)	1.38 (1.59)	4.69 (5.32)	0.50 (0.47)
Chi-square test	4.20		4.40		7.51	
P-value	0.521		0.493		0.186	

**** 1% statistical significance; ** 5% statistical significance; and * 10% statistical significance. SE=standard error.*

The corresponding regressions at the follow-up yield different results. Gender and age are no longer strong predictors of missingness, and age in particular appears unrelated to the probability of missing data. The association between missingness and EYPP and EAL remains strong, although coefficients are not always statistically significant. This suggests that, over time, the sample becomes less representative of EYPP and EAL, who are less likely to complete follow-up assessments.

Table 30: Determinants of missingness in the intervention and control groups (follow-up)

	REV		RAPT		GAPS test	
	Control group N=778[a] Coeff. (SE)	Intervention group N=810[a] Coeff. (SE)	Control group N=785[a] Coeff. (SE)	Intervention group N=839[a] Coeff. (SE)	Control group N=717[a] Coeff. (SE)	Intervention group N=777[a] Coeff. (SE)
Gender	0.72 (0.18)	0.61** (0.15)	0.91 (0.24)	0.48** (0.13)	0.76 (0.19)	0.74 (0.18)
Age	0.98 (0.03)	0.96 (0.03)	0.99 (0.03)	0.98 (0.03)	0.97 (0.03)	0.96 (0.03)
EYPP	0.90 (0.30)	2.24* (0.64)	1.04 (0.36)	2.03** (0.62)	1.00 (0.34)	2.19** (0.63)
EAL	1.70** (0.45)	1.90* (0.49)	2.25** (0.62)	1.99** (0.54)	1.71** (0.46)	1.70** (0.44)
IDACI score	10.10 (15.29)	1.22 (1.61)	6.84 (10.97)	1.08 (1.52)	3.38 (5.01)	0.88 (1.16)
Chi-square test	5.63		5.41		3.46	
P-value	0.344		0.368		0.630	

**** 1% statistical significance; ** 5% statistical significance; and * 10% statistical significance.* [a] The number of observations used in the logistic regression differ slightly from the samples illustrated in the participation flow diagram because EYPP and EAL status is missing for some observations. SE=standard error.

Tables 29 and 30 also report tests of equality of coefficients between the intervention and control groups. These tests assess whether the determinants of missingness operate differently across groups, which would imply diverging sample composition and reduced comparability. The results show no statistically significant differences between groups: coefficients are very similar in the intervention and control samples. This indicates that the composition of the two groups remains broadly comparable over time.

Overall, these results are reassuring with respect to the impact of missing data on the analysis. While missingness is correlated with characteristics associated with the outcomes and treatment responsiveness, namely, EYPP and EAL status, there is no evidence that missingness differentially affects the intervention and control groups. The proportion of missing cases is similar across groups at both baseline and follow-up, and the determinants of missingness operate in similar ways in the two groups. We therefore, tentatively conclude that missingness can be treated as occurring at random, conditional on observed characteristics. Under this assumption, more complex modelling approaches are not required. In the 'Additional analyses and robustness checks' section below, we nonetheless report results from saturated regression models that control for key predictors of missingness, including EYPP and EAL status.

Additional analyses and robustness checks

Table 31 reports the impact of the intervention on the primary outcome separately for PVIs and state-maintained settings. Although this analysis was not pre-specified in the evaluation protocol or the SAP (Masset and Butt, 2025), it was undertaken on an exploratory basis, because PVI and state-maintained nurseries differ in important ways and the IPE highlighted notable differences in implementation and perceived impact.

There were 26 PVIs in our sample. In these settings the estimated intervention effect is close to zero (Hedges' g = -0.01 [-0.20, 0.19], p = 0.927), suggesting that the overall positive effect of Talking Time© is largely driven by state-maintained settings. In contrast, the estimated effect in state-maintained settings corresponded to approximately two months' additional progress (Hedges' g = 0.10 [0.01, 0.19], p = 0.024).

Table 31: Impact of Talking Time© on the primary outcome

Outcome	Unadjusted means				Effect size		
	Intervention group		Control group				
	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n	Hedges' g (95% CI)	P-value
PVI: Renfrew composite index	127 (0)	1.02 (0.84,1.20)	122 (0)	0.99 (0.77,1.07)	249	-0.01 (-0.20, 0.19)	0.927
State-maintained: Renfrew composite index	598 (0)	1.18 (1.11,1.26)	592 (0)	1.00 (0.92,1.08)	1,190	0.10 (0.01,0.19)	0.024

Table 32 reports the estimated effects of the intervention for the five assessments under alternative model specifications. We consider five specifications in addition to the model employed in the primary and secondary analysis (a multilevel model with a random intercept):

- an OLS model with strata fixed effects and clustered standard errors;
- a multilevel model with random intercepts and slopes;
- a saturated OLS model;

- a saturated multilevel model with random intercepts; and
- a saturated multilevel model with random intercepts and slopes.

The saturated models include the following covariates: gender; child's age in months; EAL status; EYPP status; and the school-specific IDACI deprivation index. The missing data analysis showed that these variables are correlated with missingness, making it important to assess whether their inclusion affects the estimated intervention effects.

Across all specifications, estimated effect sizes and standard errors are very similar to those obtained using our preferred model (the multilevel random intercept model), indicating that the results are robust to alternative modelling choices. With respect to the multilevel random intercept and random slope model, we conducted a likelihood ratio test to assess the model fit. In all cases, the test rejected the inclusion of random slopes in favour of the more parsimonious random intercept model.

Table 32: Impact estimates under different model specifications

	Renfrew index Hedges' g (95% CI)	REV Hedges' g (95% CI)	RAPT information (vocabulary knowledge) Hedges' g (95% CI)	RAPT grammar Hedges' g (95% CI)	GAPS test Hedges' g (95% CI)
Primary analysis model Multilevel random intercept	**0.083 (0.001, 0.165)**	**0.079 (0.002,0.157)**	**0.064 (-0.014,0.142)**	**0.092 (-0.020, 0.203)**	**-0.054 (-0.136, 0.028)**
OLS with strata fixed effects	0.080 (-0.005,0.165)	0.073 (-0.006,0.153)	0.057 (-0.057, 0.172)	0.091 (-0.026,0.208)	-0.051 (-0.133,0.031)
Multilevel random intercept and random slope	0.082 (0.001,0.164)	0.080 (0.002,0.158)	0.064 (-0.047,0.174)	0.092 (-0.021,0.204)	-0.054 (-0.137,0.028)
Saturated OLS with strata fixed effects	0.082 (-0.002,0.167)	0.077 (0.001,0.152)	0.059 (-0.050,0.169)	0.096 (-0.017,0.208)	-0.042 (-0.120,0.037)
Saturated multilevel random intercept	0.084 (0.005,0.163)	0.083 (0.007,0.159)	0.065 (-0.042,0.173)	0.096 (-0.010,0.201)	-0.043 (-0.123,0.037)
Saturated multilevel random intercept and random slope	0.083 (0.004,0.162)	0.083 (0.007,0.158)	0.064 (-0.044,0.172)	0.095 (-0.010,0.200)	-0.043 (-0.124,0.037)

The administration of the GAPS test at baseline presented particular challenges. In the analysis sample for the GAPS test (1,369 observations), 50 children (3.4%) recorded zero scores at follow-up, compared with 582 children (42.5%) at baseline. There are two main reasons for the large number of zero scores at baseline. First, the assessment is designed for children aged three years and four months or older. The average age at baseline was 44 months, with approximately 25% of children aged three and four months or younger. Second, at baseline, assessors followed a stopping rule whereby the assessment was discontinued if a child failed to provide three consecutive correct responses. Around 42% of zero baseline scores resulted from the application of this rule.

There is no evidence that the GAPS test score was otherwise affected by measurement. Cronbach alpha for internal consistency was 0.96 at baseline and 0.86 at follow-up. Item-rest correlations were uniformly high, indicating that all items contribute strongly to the underlying construct. Overall, the measure appears to be reliable.

To assess the robustness of the GAPS test results we conducted two additional analyses. First, we estimated the ANCOVA multilevel random intercept model using baseline Renfrew index scores instead of baseline GAPS test score as the covariate. The Renfrew assessments were not affected by floor effects, and follow-up GAPS test scores also showed no such issues. Second, we implemented a multiple imputation approach, in which baseline GAPS test zero scores, whether

due to no completion or to the application of the stopping rule, were treated as missing. Baseline scores were then imputed using scores from all other assessments as predictors. The results, reported in Table 33, shows that both effect sizes and standard errors are very similar under these alternative specifications. We therefore, conclude that the originally reported GAPS test results are robust.

Table 33: Impact of Talking Time© on the GAPS test assessment under different model specifications

	GAPS test
Primary analysis model Multilevel random intercept (SE)	**-0.21 (0.16)**
Model with Renfrew test scores (SE)	-0.23 (0.16)
Multiple imputation model (SE)	-019 (0.22)

SE=standard error.

Estimation of ICC

Table 34 shows the ICCs for all the assessments at baseline and follow-up. The baseline sample includes the observations with completed assessments at baseline. Similarly, the follow-up sample includes all observations with completed assessments at follow-up. The ICCs are derived from the estimation of a multilevel model with random intercept and stratification variables. The right panel of Table 34 also shows the ICCs obtained from models including baseline covariates (child's age, gender, EYPP status, and EAL status).

Column-wise the ICCs are similar across the four tests, indicating that the degree of clustering is consistent across different outcome measures. This suggests that contextual factors (such as nursery characteristics) affect the outcomes in a similar way, regardless of the specific test used.

The ICCs are smaller at follow-up than at baseline. This implies a greater within-cluster heterogeneity at follow-up. Possibly, as children become older individual-level factors become more important in determining the outcomes relatively to setting-level factors.

Finally, the ICCs are considerably lower when calculated from models including baseline covariates. Adding covariates explain part of the between-cluster variance. This in turn suggests that some of the between-cluster variance is explained by differences in the composition of settings (in terms of children with different characteristics) across clusters, rather than by setting-level characteristics.

Table 34: ICCs for all assessments at baseline and follow-up

	Model with strata		Model with strata and covariates	
	Baseline (SE)	Follow-up (SE)	Baseline (SE)	Follow-up (SE)
Renfrew index	0.12 (0.02)	0.11 (0.02)	0.07 (0.02)	0.09 (0.02)
Expressive vocabulary	0.10 (0.02)	0.10 (0.02)	0.05 (0.02)	0.06 (0.02)
RAPT information (vocabulary knowledge)	0.13 (0.02)	0.12 (0.02)	0.08 (0.02)	0.10 (0.02)
RAPT grammar	0.12 (0.02)	0.09 (0.02)	0.09 (0.02)	0.05 (0.02)
GAPS test	0.12 (0.02)	0.08 (0.02)	0.09 (0.02)	0.04 (0.02)

SE=standard error.

IPE results

The results of the IPE are first introduced through a summary of the findings including how the logic model is supported. This is followed by an expanded exploration of findings that includes references to survey and interview data.

IPE results summary

Fidelity

- Fidelity to Talking Time© was generally strong across core components.
- Adherence to the timetable and sequencing was high, although staffing constraints often required rescheduling, particularly in school-based nurseries. Some practitioners adjusted the time spent on books in response to children's engagement.
- Settings were successful delivering two sessions per week, with almost all (98%) reporting achieving this most or every week. However, this was also the aspect considered most challenging, given patterns of child and staff absence. Adaptations or catch-up sessions were sometimes required, with mentors reporting that this occurred most frequently in settings with a strong commitment to the programme. Delivery in small, mixed-ability groups was the core principle of the programme that settings found the easiest to adhere to. Groups were typically arranged based on children's confidence, language level, and attendance, with occasional adjustments when group balance was affected.
- Session content was usually delivered as intended, with tailoring based on handbook prompts in line with the programmes flexible approach. Use of written plans varied and reflected increasing practitioner confidence. A minority of settings adapted activities beyond programme guidance by omitting elements or delivering them in a more adult-led way.
- Reinforcement activities were widely implemented, with most settings embedding materials linked to stories from small-group sessions into continuous provision to support ongoing talk and role-play. While approaches varied in scale due to available resources, overall practice aligned well with programme guidance.
- Trainer-mentors reported that qualitative engagement varied across professional development elements and was strongest for the in-person and online mentoring. Practitioner confidence with reflective practice varied. Practitioners struggled to schedule formal reflection sessions but were able to engage in informal discussion whenever the opportunity arose, for example, at lunchtime.
- Most settings planned to continue using Talking Time©, often with adaptations to ensure sustainability, such as adjusting sequencing, embedding Word Play into continuous provision and updating Hexagons resources. Practitioners generally felt confident to continue independently using the handbook. Generally, these adaptations retained the programmes core components and reflected the programme's aim for practitioners to tailor delivery to their children's needs.

Context and moderators

- Implementation of Talking Time© was shaped by a range of contextual factors and moderators across settings.
- Key facilitators to the delivery of Talking Time© sessions and professional development activities included practitioner enthusiasm, strong organisational planning, and existing child-led practice.
- Senior leadership support was important, particularly where time was allocated for training and reflection.
- Practitioners viewed personalised support from trainer-mentors as pivotal in shaping their confidence to deliver Talking Time© in a way that was responsive to their setting's context.
- Staffing pressures, including staff absences, limited cover, and mid-year staff changes, were the most common barriers to the delivery of Talking Time© sessions. Starting the programme in January and busy curricula added to

the challenges. Limited quiet space meant some sessions took place in noisier environments, and preparing resources added workload. Despite this, setting's delivered Talking Time© with high levels of adherence to core components.

- The delivery of professional development activities faced similar constraints. Staffing affected in-person mentoring opportunities, and time limitations restricted regularly scheduled reflection. Some practitioners required extra support to shift from adult-led to child-led approaches.
- Team working varied, with informal collaborative planning commonplace but affected by staffing and timetabling pressures. Children were generally reported as engaging well with Story Conversations and Hexagons. Some children with SEND struggled to sustain attention, while children with EAL became more confident over time.

Perceived impact

- Practitioners, setting leads, and mentors consistently reported improvements in practitioner understanding, confidence, and skills in supporting early language development. Confidence with child-led techniques, especially wordless books, increased over time.
- Programme strategies extended beyond formal sessions. Story Conversations became widely embedded in daily routines. Practitioners reported allowing more processing time, holding longer conversations, and being more intentional in interactions.
- Perceived child outcomes were strongly positive. Improvements were noted in vocabulary, confidence, turn taking, articulation, and sentence structure. Children who were quieter, less confident, or with weaker language skills were seen as benefiting most. Children with EAL were frequently described as making notable progress.
- Unintended consequences were mainly positive, including improved behaviour, strong peer modelling in mixed-ability groups, and earlier identification of language needs. Some younger children in open-plan PVIs benefited indirectly from observing sessions. One practitioner noted that removing staff from continuous provision at times made behaviour management more difficult.

Usual practice

- Oral language development was already prioritised across intervention settings. Schools and maintained nurseries relied more on whole-class story reading and vocabulary reinforcement, while PVIs used more child-led, small-group storytelling. Staff confidence varied, particularly among less experienced staff.
- Many settings had previous experience with oral language focused interventions.
- Talking Time© was viewed as more structured and intentional compared with usual practice. Schools highlighted the shift from reading stories to children to facilitating conversations in small mixed-ability groups based on the pictures in storybooks. PVIs felt the programme made existing practice more deliberate and provided clearer strategies.
- Control settings also prioritised language development, using a range of approaches. Some adopted elements similar to Talking Time© unintentionally, while others avoided changes to maintain control group status.
- Interpreting results must consider that Talking Time© built on existing practice rather than replacing it, and that starting points varied. Schools tended to be more adult-led, while PVIs were more familiar with child-led approaches. The use of similar approaches in some control settings may reduce contrast between groups.

Support for the programme logic model

- Findings strongly support the logic model assumptions that professional development would improve practitioner knowledge, confidence, and skills, leading to enhanced language-supporting practice. Practitioners reported

increased confidence with child-led approaches, more intentional interaction and widespread embedding of activities like Story Conversations.

- Practitioners strongly perceived positive short-term outcomes for children, including improvements in confidence, vocabulary, narrative skills, and turn taking. These gains were most pronounced among quieter children, those with weaker language skills and children with EAL.
- Exposure to high-quality practice around oral language was reported widely however, difficulty maintaining consistent dosage due to absences and staffing pressures may have affected the strength and consistency of children's experience of the intervention.

Fidelity

To understand the extent of fidelity during implementation, we explored whether the programme was delivered as intended and if any adaptations to the programme took place during the course of delivery. Flexibility is a key component of Talking Time©, with settings encouraged to adapt delivery to meet children's needs. However, there are a number of core components that settings are instructed to retain to ensure delivery adheres to the programme's key principles. These include: delivery in small mixed-ability groups of no more than five children; sessions delivered twice a week for 15 minutes; and the inclusion of reinforcement activities into continuous provision.

During the 20-week structured period, settings are also encouraged to follow the provided timetable. Books have been carefully selected to ensure structured progression across the programme, and activities are ordered to align with professional development content. However, unlike the core components, programme guidance recognises that adaptations to the timetable may be necessary to accommodate staff or child absences.

IPE RQ1: To what extent were the programme's core components implemented as intended?

Timetable/sequencing

The timetable provided in the Talking Time© handbook outlines 20 weeks delivery of two 15-minute, small-group sessions a week. The programme is segmented by six different books which are set in a specific order. In order of delivery, the books provided by the programme are: Goodnight Gorilla; Lulu Loves Flowers; Peace at Last; Chapatti Moon; and Bee and Me. The first book is delivered over four weeks, while the following books are assigned three weeks each. In the final four weeks of the programme, settings are expected to deliver the programme using a book of their choice. The first activity delivered within the 20-week programme is Story Conversations. From Week 4, Word Play is introduced as the second activity of the week until practitioners are advised to move this into continuous provision from Week 12. From this point onwards, the Hexagons activity replaces Word Play as the second activity of the week. Following the end of the 20-week period, settings are encouraged to continue delivery of the programme using their own materials.

Across interviews, practitioners in all setting types generally reported good adherence to the prescribed timetable outlined in the handbook, while noting the need for occasional adjustments in response to staffing constraints and children's engagement. Good levels of adherence are reflected in the survey data, with over 98% of respondents (n=51) reporting that they stuck to the timetable 'most weeks' or 'every week'.

Table 35: Practitioner self-reported adherence to timetable

Adherence	Never % (n)	Very few weeks % (n)	Some weeks % (n)	Most weeks % (n)	Every week % (n)
Adherence to the timetable set in the handbook	0% (0)	0% (0)	1.9% (1)	44.2% (23)	53.9% (28)

In interviews, staffing constraints emerged as a particular issue for timetable adherence in school-based nursery settings, with practitioners explaining how insufficient staffing numbers led to the rescheduling of sessions. As a result, it was not always possible to fit two sessions into a week.

Some school-based nursery settings expressed frustration around the inflexibility of the sequence of books and activities during the structured 20-week period. Practitioners felt greater flexibility in how books were ordered would have helped align Talking Time© activities with wider curriculum themes, for example, one practitioner commented that it would have been beneficial to schedule Lulu Loves Flowers, a book about plants and flowers, during the time in the year they focused on outdoors and gardening.

> *I think, it would have been nice for us to choose which book we were doing when, just because we had a whole garden theme, but we were doing Peace at Last while that was going on. It would have been nice if I could have looked at it and thought, 'Do you know what? We're going to do Lulu's Flowers linking to this topic'.* (TP16, School-based practitioner)

On the other hand, practitioners in PVI settings noted that the books mapped on to their curriculum well, potentially indicating the different levels of flexibility setting types have. Outside of concerns around alignment with the curriculum, active adjustments were made to the timetable and sequencing of Talking Time© activities based on children's engagement. Most of these adjustments remained in line with the programme's flexible approach, although highlight the extent to which *engagement* rather than children's absences or staffing issues also influence adaptations. This was most likely to be reported by practitioners in PVI settings who are arguably more used to curriculum flexibility. One practitioner from a PVI setting explained in their interview that they extended delivery of Lulu Loves Flowers for an extra week and shortened the time spent delivering Chapati Moon based on children's level of engagement with each book. One practitioner described making an adaptation to a core component of the programme in response to levels of engagement. They described delivering longer sessions (30 minutes or more) for Story Conversations and Word Play to allow children more time to settle into the activities. Talking Time© guidance suggests a maximum activity time of 15 minutes. Despite these adaptations being more prevalent in PVI settings, one school-based practitioner also reported making changes to timetable sequencing to better support children's engagement. They found that in their setting, children retained stories better when two Story Conversations sessions were delivered in one week followed by two Hexagons sessions the next week.

Trainer-mentors confirmed that while most settings maintained adherence to the timetable and sequencing, adaptations that aligned with the programmes flexible approach took place in some settings, including delivering Word Play before Story Conversations within a delivery week to provide children with exposure to activities before they were addressed in books, which was considered particularly helpful for children with additional needs and for sustaining engagement in Story Conversations.

Frequency

A core component of the Talking Time© intervention is that children received two sessions of timetabled activities a week.

Table 36: Practitioner self-reported adherence to frequency

Frequency	Never % (n)	Very few weeks % (n)	Some weeks % (n)	Most weeks % (n)	Every week % (n)
Ran two sessions per week	0% (0)	0% (0)	1.9% (1)	32.7 (17)	65.4% (34)

Survey data shows that there was strong adherence to session frequency, with 98.1% (n=51) of respondents reporting that they adhered to the dosage of two sessions a week 'most weeks' or 'every week'.

Despite this strong adherence, the interview data showed that meeting the prescribed dosage of two Talking Time© sessions per week was the most challenging core component of the programme to meet. Practitioners and trainer-mentors identified children's absences as the most significant issue to manage across all settings. School and maintained nursery school practitioners also highlighted that staff capacity and clashes with school calendar events prevented the two-dosage requirement being met some weeks. School-level barriers were considered easier to mitigate than children's absences because sessions could be rescheduled for all participating children. To make up for missed sessions, settings commonly ran 'mop-up' sessions at the end of the week or began subsequent sessions with brief catch-ups for children who had been absent prior. This was described by a practitioner from a school-based nursery:

> *We did a quick catch-up at the beginning of the next one, just so that they could at least see the bit that they'd missed, and that worked most of the times.* (TP03, School-based practitioner)

In some cases, the level of absence of an individual child meant the required frequency could not be achieved by mitigation strategies.

PVI practitioners highlighted attendance patterns as a particular challenge, citing the varied hours children attended provision as presenting an obstacle to scheduling sessions at times all participating children could attend. A senior leader interview who was also a Talking Time© lead described organising a delivery schedule around morning and afternoon nursery attendees as *'a logistical nightmare'* (SL05, PVI practitioner). One practitioner also explained how she felt that parents/carers perceived attendance at her setting as optional, which she did not believe would be the case if their children attended a school nursery.

Trainer-mentors concurred that children's absences, staff absences, and calendar clashes were the main obstacles, and observed that while settings generally adhered to core components, the twice-weekly dosage was the element least likely to be met. They also noted that practitioners' commitment to the programme influenced whether catch-up strategies were implemented. Highly committed practitioners often invested additional time to ensure children caught up on missed sessions, whereas settings with less committed staff typically had no catch-up strategy in place.

Small mixed-ability groups

Further core components to the Talking Time© intervention are universal and small-group delivery. In practice, this required settings to convene small mixed-ability groups of up to five children to deliver activities to.

Table 37: Practitioner self-reported adherence to small mixed-ability groups

Adherence	Never % (n)	Very few weeks % (n)	Some weeks % (n)	Most weeks % (n)	Every week % (n)
Used small mixed-ability groups of five children or fewer	0% (0)	0% (0)	0% (0)	25.0% (13)	75.0% (39)

Adherence to this core component according to self-reporting in surveys was the strongest with all respondents reporting they delivered to small mixed-ability groups 'most weeks' or 'every week'. This was evident in interviews with all practitioners. Across setting types, good adherence to small mixed-ability groups were reported. Alongside language ability, practitioners used children's confidence levels and EAL status as key determinants for group composition. PVI settings also cited children's attendance patterns as key in deciding on groups, with a few also including some of their high ability two-year-olds in Talking Time© groups, despite being asked to include three to four-year-olds only (the programme is designed for three to five-year-olds). While efforts were made to maintain consistent groups throughout the delivery of the intervention, adjustments were occasionally necessary to accommodate absences or if during the course of the intervention it was felt there were too many confident children concentrated in one group. Delivery in small mixed-ability groups was widely regarded as beneficial, with practitioners emphasising the value of confident children modelling language while highlighting the need to manage participation so that quieter children were not overshadowed. As one practitioner explained:

> *I think it benefited the children a lot because you have the chatty ones, and then the others try to copy and mirror their behaviourisms, or the communication and language. However, it can be a little bit tricky sometimes to try and pause the chatty ones, because then if you need to give some space to other children to join in in a conversation.* (TP19, School-based practitioner)

Session content

In interviews, practitioners commonly described adapting session content as recommended by the Talking Time© programme, namely, adjusting prompts and plans so they suit the language level of the different children in each group. Survey data shows there is variable use of plans and prompts across sessions, which is to be expected as the programme's aims for practitioners to become more self-sufficient as the programme progresses and to use their knowledge of their children to plan sessions with necessary adaptations.

Table 38: Practitioner self-reported use of plans and prompts from handbooks

	Never % (n)	Very few sessions % (n)	Some sessions % (n)	Most sessions % (n)	Every session % (n)
Frequency of use of plans and prompts provided	0% (0)	5.8% (3)	26.9% (14)	36.5% (19)	30.8% (16)

The different levels of prompts and questions provided in the handbook were described in interviews as key tools for tailoring sessions to the appropriate level for children. However, some PVI practitioners reported adjustments that were less consistent with the programme's principles, such as omitting activities due to poor engagement or delivering activities in a more adult-led way than intended. For example, one setting excluded Chapatti Moon Word Play activities from continuous provision because children showed limited interest, while another included two-year-olds in Word Play activities once these were placed in continuous provision. This required more adult support than recommended to manage younger children's participation.

Reinforcement activities

Practitioner implementation of reinforcement activities (i.e. adding an activity into provision linked to small-group activities) was widespread across the interviewed settings. Most practitioners described placing role-play materials, often drawn from Word Play sessions, into areas of continuous provision so that children could continue conversations initiated during small-group activities independently.

> *We did an awful lot around Play-Doh and making chapattis with Play-Doh when we did Chapatti Moon [...] we did it as an activity in provision, and the children were doing it themselves afterwards [...] We would look at what we'd done and add it into our planning and then do it. It might have been one or two days a week, or maybe just adding them in and not even really talking about it, just putting it there and seeing what they did with it.* (TP04, Maintained nursery practitioner)

A few practitioners described more immersive approaches, which included trips or practical activities, with one practitioner outlining a structured set of experiences over the course of a book's delivery period:

> When we did, *Lulu Loves Flowers*, we did a trip to the library, we did a trip to go and buy seeds on that trip. Doing all of those things that Lulu did at the same time was really, really powerful for them [...] Then we had a picnic like she did at the end of the story. We did that over a week and a half. So, making those experiences real from the story, and then linking it back to a story was really powerful for their learning. (TP17, School-based practitioner)

Across settings, these reinforcement activities were reported to be well engaged with by children, who were frequently observed using vocabulary from sessions and independently role-playing story content during their play.

Some practitioners noted that planning and refreshing activities within provision could be time-consuming. However, the Talking Time© handbook was consistently referenced as a useful source of practical ideas, particularly when time or inspiration was limited.

> *You know, so sometimes, it might be a tuff tray with, say, from the gorilla story at the beginning, we made a little zoo, and that would be out but sometimes it was difficult to change that every time because, obviously, we're quite busy and we have a whole room to change. […] I will say the handbook was really helpful to have because sometimes, if you're stuck for an idea, you're trying to think of something but you're like, 'Actually, I'll just go to the handbook and see what they have in there.'* (TP07, Maintained nursery practitioner)

Overall, this pattern of implementation indicates good fidelity to the intended principles of reinforcement activities, with settings embedding story-linked activities into continuous provision in ways that aligned with programme guidance. The scale of this delivery varied across settings based on available resources.

IPE RQ2: Is the intervention feasible for practitioners to engage in and implement as intended? Does this vary by setting type?

Based on interview discussions, practitioners generally considered the intervention feasible to engage with and implement as intended. Additionally, survey data shows good adherence across the core components of the programme.

All those interviewed, described being able to embed sessions into their timetables, and in principle, the required dosage was achievable. However, the most significant challenge to meeting this dosage was children's absences, which was beyond practitioners' control. This issue was reported across all setting types, with PVI practitioners highlighting that varied attendance patterns posed an additional hurdle to maintaining key components of the programme. While strategies such as rescheduling and catch-up sessions were used, high levels of absence sometimes made it difficult to achieve the intended frequency.

IPE RQ3: How do practitioners engage with the training and the follow-up mentoring? Which practitioners are selected to attend training and deliver the programme (including how/why)? 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?

Trainer-mentor reflections on engagement with professional development activities

The professional development element of Talking Time© comprises four modes of support: twilight sessions; in-person mentoring; online mentoring; and self-guided reflection. Twilight sessions involve clusters of settings attending online workshops delivered by trainer-mentors after school hours to introduce activities and strategies. For those who were unable to attend, recordings of these sessions were sent to Talking Time© leads. In-person mentoring provides opportunities for trainer-mentors to model activities and for practitioners to practice these with feedback. Online mentoring requires practitioners to submit video recordings of their delivery, which trainer-mentors support practitioners to analyse and identify strengths and areas for development. Self-guided reflection is facilitated by resources in the Talking Time© handbook and intended to take place between the two Talking Time© leads.

Interviews with trainer-mentors offered valuable insights into practitioner engagement across these different areas of training and mentoring. Engagement with twilight sessions was considered mixed. Those who attended participated well, but attendance was challenging for some practitioners due to sessions being scheduled outside of contracted hours, creating barriers for practitioners with other commitments. In-person mentoring was regarded as particularly useful for practitioner engagement in in-depth and nuanced discussions about how Talking Time© strategies differed from usual practice, although it was noted that these sessions were disrupted at times by competing demands within settings.

> *With the in-class mentoring, I think the real power of that is that they see how this programme works with their children. So, when they have that, sort of, 'Oh, this is the kind of thing we do anyway,' or sometimes*

> *they get that, 'Oh, that's not going to work for our kids,' seeing what that looks like in practise with their children, I think is incredibly powerful. And having the opportunity to have a conversation about that in some depth, I think is also really powerful [...] it really helps us to understand the levels of adaptation that are needed both for the children and for the staff. I think it would be difficult to do that to that level of quality in another way. The downside of that is that stuff happens in nurseries. Children bump their heads, you know, staff shortages. So, sometimes in some situations, the staff were a bit distracted or somebody was pulled away to do something else."* (TM04)

Online mentoring was perceived by trainer-mentors as providing practitioners with the rare opportunity for quiet and private space to engage in focused reflection and open dialogue about their practice away from the fast-paced environment of provision.

> *The online mentoring was a fabulous platform for mentoring because they would find a quiet room, to have the online meeting. So, they were out of it, so, they had lots of space and head space and we just had some very, very good then discussions in that scenario*. (TM03)

The programme handbook provides a reflective tool, which Talking Time© practitioners are guided to write in or use a guide for reflective discussions with colleagues. Although not a compulsory element to professional development activities, written reflections were reported by trainer-mentors as the mode of professional development that was engaged with least. This was considered to be due to time pressures. However, reflections during trainer-mentor visits were seen as working well to support practice throughout the 20-week period of delivery.

> I didn't have any of the settings that did written reflections. [...] They were making refinements to practice, because of what they had observed their colleagues do during the session. So, that was working really, really well, but the written part of the reflections, none of them did, because none of them had the time to do it, and possibly even the confidence to do it. (TM02)

The lack of time for written or scheduled reflection sessions was echoed by practitioners interviewed, who described challenges finding time to timetable formal reflection sessions between Talking Time© leads and complete the corresponding sections of the handbook. In schools, this was often compounded by the fact that the second Talking Time© lead was a teaching assistant, roles which are generally only contracted during provision hours. This left little opportunity for scheduled reflection outside of teaching hours. The programme guidance acknowledges the likelihood of these time constraints and encourages reflection in any way possible, even if for a few minutes. Across all setting types, practitioners commonly spoke about relying on regular but informal moments to reflect on sessions and professional development, such as quick conversations over lunch or just before the children arrived in the morning.

There were conflicting opinions about how practitioner experience level and setting type influenced engagement with reflective practices. Some trainer-mentors felt that less experienced practitioners struggled to see the value of reflective practices because they were less familiar with using them, and that confidence building and reinforcing basic strategies was more important when working with these practitioners. In contrast, one trainer-mentor commented that in their settings it was the more experienced practitioners who found noticing and reflecting more challenging than junior staff, as they found it difficult to adjust to an approach that was centred on guided reflection rather than feedback. Two extracts from interviews below demonstrate these contrasting points:

> *So if I think about the stronger practitioners, it felt more like we can have an open discussion, we can have a, kind of, true reflection, whereas for the weaker practitioners I found myself going back to the manual and saying, rather than saying, 'So how did you feel about planning this?' I knew she hadn't planned because from the session it was obvious, so rather than putting her on the spot and saying, 'What did your plan look like?' I said, 'So let's go back to these pages of the manual that support your planning,' and just more information sharing about the basics*. (TM07)

> *I had to demystify the idea that I'm not here for [telling them what they are doing right or wrong]. I'm here to guide you to reflect. So, I would say that the first in-class mentoring for all of them was a little bit difficult [...] the more experienced the practitioner, the more difficult it was for them.* (TM02)

Regarding setting type, a few trainer-mentors observed that school-based practitioners tended to be more 'negative' towards reflection and struggled to engage with it, as they were more used to external observations that resulted in feedback that felt like criticism rather than support. As one trainer-mentor explained:

> *The school staff in general, I found to be more negative. Not in terms of they weren't able, but I think they were so used to getting criticism [...] [teachers are] just constantly observed and then it's like, nit-picking at things. And I think they went into all of those sessions with that mindset of, 'Well what have I done wrong? What do I need to do better?' [...] They wanted a lot of external feedback and they weren't used to having to reflect on their own processes and their own practice really. I think they were so used to having all of that externally that they weren't necessarily engaging in that process themselves on a regular basis. So, it surprised me in a way, because you, kind of, think that school staff would be more, and it's not that they weren't able to listen to it but they just, they were quite negative in the sense of, 'Oh, you know, what are you going to hit me with now?'* (TM05)

Another trainer-mentor, however, felt that practitioners in PVI settings struggled more than those in schools because they were less accustomed to engaging in reflective practice.

Ultimately, these observations suggest that individual openness to new ideas and to having practice critiqued are key factors in facilitating engagement with reflective practices.

Talking Time© lead selection process

Across schools, setting leads (including headteachers, assistant heads, and Early Years Foundation Stage [EYFS] leads) described selecting Talking Time© leads based on developmental needs. For example, one lead spoke about choosing an early career teacher and a learning support assistant, who they felt could benefit from skills development and space to reflect on practice. Other leads explained that they chose staff who would be strongest at disseminating professional learnings to the wider team.

In PVI settings, the first Talking Time© lead was often the nursery manager, who typically self-selected as the most qualified early years practitioner, often with decades of experience. Managers were also more likely to be paid for administration time, enabling full engagement with professional development activities. A key advantage of this arrangement was that managers were generally not in continuous provision full-time, so taking groups aside for Talking Time© caused minimal disruption to staffing. Some PVI leads highlighted concerns about time constraints for junior staff when selecting the second lead or, in the few cases when the manager was not a Talking Time© lead, when selecting both practitioners. Reasons for selecting non-manager Talking Time© leads varied, including choosing practitioners already working with the target age group, those who spoke different languages to support wider parent/carer engagement, or staff specialising in speech and language.

Content accessibility

The accessibility of professional development content did not receive much comment in interviews with practitioners suggesting this was not an issue for most. As outlined above, issues around accessibility were highlighted to be those around the ability to engage fully in professional development sessions based on the time and location they took place. That said, one PVI setting lead (who was also a Talking Time© lead) did mention that they felt that the handbook read as those it was written for school practitioners who were of qualified teacher status and above. They believed at this level, practitioners would be familiar with some of the more theoretical concepts. They commented on how some of their more junior staff who were Level 3 or below, would struggle to find the handbook information useful as it was set at too high a level. This had implications for their plans to train the wider staff body to deliver Talking Time© sessions. In contrast, one

practitioner we did speak to who had a Level 3 qualification, considered the content to be in line with practice they were already familiar with, and they engaged with the content well.

Interviews with trainer-mentors were more illuminating for understanding the varied extent to which practitioners were able to access the content of Talking Time© professional development activities. Most notably, trainer-mentors reported how several practitioners struggled with grasping a child-led approach and the main aim of an activity. A key enabler to Talking Time© content being accessible to a range of early years practitioners were trainer-mentors' strategies for adapting their practice and delivery styles to meet practitioners' needs and ability. They described different ways they did this including using more direct methods of critique for those struggling to transition to child-led approaches; dedicating online mentoring sessions to revisiting activities (often the purpose of Word Play) rather than using the time for reflection when practitioners struggled to engage children with high-quality conversations; and adjusting the level of delivery, namely, the amount they engaged with theoretical concepts, based on the practitioners' knowledge level.

Types of training and support valued from trainer-mentors

Practitioners across all setting types consistently emphasised in interviews that they highly valued trainer-mentors' responsiveness. Timely responses to queries via email or phone helped them feel supported and stay on track with Talking Time© timelines.

> *It was just fantastic support [...] if I've got a question or anything, I ping off an email and [our trainer-mentor] gets straight back to me. [...] So, that was a brilliant two-way teamwork to deliver it. And I know that won't always continue [...] but that was more useful to us, almost, than the strategy.* (TP20, School-based practitioner)

In-person mentoring was viewed as particularly important for mentors to understand contextual factors such as intake demographics and operational demands. These visits allowed mentors to experience delivery first-hand, enabling them to provide tailored support and suggest feasible strategies.

Several school-based nursery practitioners also commented on appreciating mentors' professional experience in early years practice, which made guidance practical and relevant. Support was perceived as empathetic and grounded in a first-hand experience of delivering in early years settings.

> *I felt totally supported by [our trainer-mentor], she's a lovely person and she put me at ease, as well. And, because of her experiences, and the fact that she is passionate about Early Years, unless you are passionate about Early Years and child development and understand child development, this programme wouldn't work. Because you've got to have that ground knowledge and she really did.* (TP11, School-based practitioner)

PVI practitioners particularly highlighted the importance of mentors' efforts to build their confidence through consistent encouragement and carefully tailored individual feedback, which they viewed as fundamental to their professional growth.

> *Our mentor is awesome, she is superb. She pays attention to details and helps us—me and my partner—she helped us to grow and know more about the profession itself. Not only about Talking Time©, but about the profession itself.* (TP05, PVI practitioner)

Additionally, practitioners in maintained nurseries seemed to especially value trainer-mentors' calming approach, noting it helped reduce feelings of being overwhelmed when completing new activities and made them feel comfortable with having their practice observed.

Support for colleagues who do not directly access training

The 20-week structured period of delivery required the two assigned Talking Time© leads to attend the twilight and in-person mentoring sessions. The main lead was also required to attend online mentoring sessions. This structure was designed for the purposes of the efficacy trial to enable support for a large amount of settings in a short space of time. Outside of the trial—the programme is designed to support all early years practitioners in a setting.

Some practitioners mentioned in interviews that they had begun to share responsibility for delivering Talking Time© with practitioners who were not selected as Talking Time© leads and therefore, had not attended the full suite of training sessions. This was often driven by a desire to increase the number of children accessing the programme.

> *I've linked [the second Talking Time© lead] to another member of staff who's a level two who isn't very confident at all. She's going to work with this young girl and then I'm going to do the same with the other three, four members of staff [...] So, hopefully, yes, moving forward from September, all the staff will be doing small group works with all the children.* (TP02, PVI practitioner)

Typically, staff members who were being trained up internally had already attended twilight sessions alongside the designated Talking Time© leads and had begun observing trained colleagues with plans for further peer training in the new year. In a few instances, these practitioners had already begun delivering to new groups of children. To support the training of new staff, settings were provided with induction videos. Despite this, some school-based practitioners noted that, while they wished to share learning with other early years staff, they lacked the capacity to do so.

IPE RQ4: 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?

Plans to continue and adaptations

The vast majority of survey respondents planned to continue using Talking Time© in some form after their participation in the trial ended. Only 11.1% (n=6) responded that they would 'maybe' continue and no respondents responded 'no' (Table 39).

Table 39: Setting plans to continue using Talking Time© approaches or activities

	% (n)
Yes, definitely	59.3% (32)
Yes, with adaptations	29.6% (16)
Maybe	11.1% (6)
No	0.0% (0)

Most settings interviewed expressed plans to continue implementing Talking Time©, though trainer-mentors noted that the strength of plans often depended on factors such as senior leadership support and staff retention, including whether trained practitioners remained in the same year group.

School setting leads commonly reported intentions to train additional staff to sustain delivery in nursery and expand the programme into Reception. Some schools planned to extend delivery even further into Year 1, with one practitioner believing the programme would be more beneficial for children with SEND in Reception and Year 1 than in nursery. PVI settings also described plans to train more staff to reach a wider cohort of children, though there was less discussion about changing the age group of delivery. Where this was considered, adaptations focused on including two year-olds in sessions by making some activities more adult-led.

As the survey data shows, a significant number of respondents intend to continue with Talking Time© with adaptations (29.6%, n=16). Plans to adapt programme structure and resources to ensure sustainability were frequently mentioned in interviews. Most changes aligned with the flexibility embedded into the programme, with core components retained. Planned changes included spending longer than three weeks on books, introducing new books that are more relatable to children's social circumstances, and reordering existing books to align with the settings curriculum plan. For example, reading Lulu Loves Flowers during planting and growing topics. These small adaptations to Story Conversations were considered necessary to improve children's engagement in sessions. One planned adaptation that was mentioned by a number of settings that disrupted one of the core components was the move of Word Play exclusively into continuous provision. This shifted it from being an activity that could be guaranteed to be in small mixed-ability groups. The description of plans going forward by the following practitioner was a common sentiment across interviews:

> *We almost want to do it exactly the same, but we don't think we would do Word Play in a little take-out group. Word Play would just be in provision, and we would start with Story Conversations and add Hexagons in later as we have done.* (TP04, Maintained nursery practitioner)

It was also common for practitioners to mention plans to create new Hexagons resources featuring images that better reflect activities that are familiar to children. This was actively encouraged during the final mentoring sessions.

While most settings included in interviews wanted to continue with Talking Time©, some uncertainty about the feasibility of this was expressed across different setting types. For example, two school-based practitioners were unsure about continuation, citing concerns that the high numbers of children with SEND in the incoming nursery cohort would make implementation difficult. This is because it was felt that removing staff from continuous provision would be too disruptive for this group of children. A practitioner in a maintained nursery school hoped the programme would continue but noted that staffing levels would be a determining factor, particularly as their class planned to merge with Reception. Regardless of continuation, they felt the Continuing Professional Development (CPD) received, especially around adopting a child-led approach to reading, would remain embedded in their practice.

> *I do think it is something that we've enjoyed being part of, and we've definitely, it's really supporting our CPD, and the way that we're looking at stories, and I think that will definitely move forwards. I'm hoping that we'll be able to do Talking Time© in some way, shape, or form, but I just don't know the details [around staffing] yet, moving forward, which will impact what's going to happen with it.* (TP12, Maintained nursery practitioner)

One PVI setting reported they would not continue Talking Time© in a structured way, viewing it as similar to their usual practice, but planned to incorporate Story Conversations and Hexagons into their activity bank and to maintain mixed-ability groups, which they felt had a positive impact on children with lower language and communication levels.

Trainer-mentors highlighted other reasons for non-continuation including: staff changes (such as trained practitioners leaving or going on maternity leave); the programme's lack of accreditation which meant for senior leaders the resources required to go into the programme could not be justified without knowing the programme's impact; and the stress the delivery of the programme induced for a few settings due to persistent implementation challenges, such as child absences, which meant they wanted to relieve themselves from the programme once the trial had ended.

Practitioner confidence to continue

Survey data shows high levels of confidence to continue without trainer-mentor support among respondents with 92.8% (n=53) rating themselves at 4 ('confident') or above (out of 5 'very confident') (Table 40).

Table 40: Practitioner self-rating of confidence to continue

	1 (Not confident at all) % (n)	2 % (n)	3 % (n)	4 % (n)	5 (Very confident) % (n)
Confidence to continue	0% (0)	0% (0)	1.9% (1)	27.8% (15)	70.4% (38)

Note: respondents were asked to rate confidence on a 5-point scale, ranging from "Not confident at all" (1) to "Very confident" (5).

All school settings interviewed expressed confidence in continuing the programme without external support, viewing the Talking Time© handbook as a sufficient tool for future implementation and training additional staff. Reflecting this sentiment, one school-based nursery practitioner commented:

> *The manual itself and the resources we were provided with, they're so thorough and give so much information and detail that I think with that alongside us, it would be relatively easy to continue the programme without the support of the mentor.* (TP03, School-based practitioner)

PVI settings also expressed confidence in continuing the programme independently, although some PVI practitioners mentioned that they would value some follow-up support from mentors to review how effectively they have embedded Taking Time© into their provision. In line with this, some trainer-mentors felt that settings could benefit from catch-up mentoring sessions to ensure the programme has been sustained effectively. In the absence of plans for this, one trainer-mentor suggested to their settings to establish communities of practice to support ongoing implementation and sharing of best practice.

Context and moderators

This section explores the range of issues affecting implementation of Talking Time©, identifying the enablers and challenges that contribute to successful implementation as understood by practitioners, senior leaders, and trainer-mentors.

IPE RQ5: What are the potential barriers and facilitators to the delivery of the programme? Do these vary by setting type?

Based on survey data, 76.4% (n=42) of respondents considered implementation to be 'manageable', 'easy' or 'very easy' to implement. However, a substantial minority (23.6%, n=13) found it to be 'difficult' or 'very difficult' (Table 41).

Table 41: Practitioner perception of manageability of implementing Talking Time©

Manageability	% (n)
Very easy	3.6% (2)
Easy	18.2 (10)
Manageable	54.6% (30)
Difficult	20.0% (11)
Very difficult	3.6% (2)

Barriers to intervention delivery

Among the reasons mentioned for the difficulties within survey responses, staff shortages and staff time pressures were predominant.

Interviews with trainer-mentors also identified staffing as a key challenge to delivery that affected all setting types. Staff absences and insufficient capacity to cover continuous provision was reported as affecting adherence to the required session frequency. Additionally, mid-year staffing changes due to sick leave or redeployment to other year groups meant some practitioners joined the programme late, impacting delivery quality.

Beyond staffing, trainer-mentors shared their observation that starting the programme in January rather than September hindered the programme's integration into practice (this was necessary to accommodate the evaluation timetable). Additionally, trainer-mentors shared their observation that overpacked curricula with multiple interventions further constrained capacity to deliver Talking Time© at high quality. Across school-based nurseries, settings without leadership buy-in were observed as struggling the most, as practitioners were left to operate independently without the additional support needed. In contrast, leadership buy-in was seldom an issue for PVI settings, as nursery managers were commonly one of the Talking Time© leads.

In interviews with practitioners across all setting types, a common challenge reported was finding additional quiet space for Talking Time© sessions. Practitioners found creative ways to navigate this. Some settings had access to spare rooms such as book rooms or spaces used for specialist interventions, while many others relied on alternative areas such as offices, empty classrooms during outdoor play, kitchens, or cloakroom areas. Practitioners in both school and PVI settings reported that on occasions, due to limited staff and space, they had to deliver sessions in the same room as continuous provision. This was described as 'hectic' and distracting for children, affecting their engagement in sessions.

Some practitioners also noted the extra planning time required for new activities as a barrier to integrating Talking Time© seamlessly into daily routines. As one practitioner from a PVI explained:

> *It's not just another activity that you're adding in, there's all the prep behind it, and if you're not constantly doing that sort of reflection and additional prep, it's quite new [...] I thought it would just slot in after a couple of months, but it didn't. It was too different.* (TP06, PVI practitioner)

Additionally, although not a formal requirement, one school practitioner noted that some activities, particularly Word Play, required purchasing or making new resources such as locks and cardboard cages for the weeks on Goodnight Gorilla, which was challenging on a limited school budget.

Barriers to professional development delivery

Trainer-mentors reported several barriers to delivering the professional development component of Talking Time© effectively in their interviews. Some barriers were related to specific modes of delivery. For in-person mentoring sessions, limited staff capacity often made it impossible for two staff members to leave provision at the same time (although practitioners were offered the option of doing it individually). For PVI settings in particular, varied attendance patterns of children made it difficult to gather enough children to conduct effective demonstrations and reflections. Trainer-mentors also noted that formal reflection sessions between staff faced barriers, primarily due to time constraints. A key part of the professional development element of Talking Time© was for Talking Time© leads to spend time every week reflecting on their practice using the 'rehearsing, noticing, analysing, and refining' tool provided in the handbook. Schools without senior leadership buy-in lacked ringfenced time for reflective practice, and differences in staff roles created unequal opportunities to engage in reflection. Teaching assistants, for example, rarely had time in their schedule to reflect alongside teachers. Additional barriers included communication gaps between Talking Time© leads, which led to missed training opportunities for second leads; the predominant use of adult-led pedagogical approaches, requiring more time with practitioners than often assigned as a more significant shift in practice than those who already incorporated child-led approaches was needed; and the need for mentors to spend considerable time motivating less experienced practitioners rather than focusing on teaching programme principles.

Facilitators of intervention and professional development delivery

Trainer-mentors described a combination of factors that facilitated both delivery of Talking Time© sessions and the professional development component. For intervention delivery, effective management of programme logistics was considered essential. Maintaining small-group sizes supported behaviour management and engagement, while consistent scheduling helped to streamline organisation of staff and children. Practitioner attitudes were also perceived to play a significant role. Strong engagement, advocacy for the programme within the curriculum, and a willingness to try new approaches were repeatedly linked to successful implementation. Setting's pedagogical approaches also played a role.

One trainer-mentor highlighted that PVIs with existing child-led approaches were the settings she worked with that were best able to integrate Talking Time© into their provision. Another trainer-mentor observed that settings where practitioners were confident in tailoring delivery to their children's needs, such as adapting resources to create relatable experiences, delivered the programme most effectively and achieved the biggest impact.

Similar factors supported professional development delivery. Trainer-mentors highlighted how practitioner enthusiasm and willingness to engage with new strategies were key enablers across all modes of training. Pedagogical flexibility again helped; settings with child-led approaches adapted quickly, allowing professional development to focus on intentionality of practice. Additionally, organisational support, particularly senior leadership buy-in, was critical for ringfencing time for reflection, which strengthened professional learning. Practical considerations also mattered, with access to quiet spaces enabling in-person and online mentoring to run smoothly.

IPE RQ6: 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?

Perceptions of resources

Across interviews with practitioners of all setting types, there was an overwhelming positive view of the Talking Time© handbook and the guidance provided, with these resources considered fundamental in supporting their adaptation of activities to align with children's needs and abilities. Some practitioners described relying heavily on the handbook throughout the programme, while others explained that they utilised it extensively at the start but later on, as intended by the programme, were able to plan out sessions without much dependence on it. No particular setting type stood out as being more or less reliant on the handbook than others.

> *I think we relied on [the handbook] much more with the first few books, because you just wanted to make sure you were getting it right. But actually once you'd done it a few times you, kind of, realised you were asking the questions that were in the planning without even really having a proper look at it. You might have had a skim through, but it almost felt like you knew what to ask because you'd done so much of the training and you'd worked through it before.* (TP04, Maintained nursey school practitioner)

Table 42: Practitioner reported helpfulness of resources

	Not at all helpful % (n)	Slightly helpful % (n)	Moderately helpful % (n)	Very helpful % (n)	Extremely helpful % (n)
Story Conversations plans and prompts	0% (0)	0% (0)	1.9% (1)	38.9% (21)	59.3% (32)
Word Play plans and prompts	0% (0)	0% (0)	5.6% (3)	44.4% (24)	50.0% (27)
Hexagons plans and prompts	0% (0)	1.2% (1)	9.3% (5)	31.5% (17)	57.4% (31)
Talking Time© handbook	0% (0)	0% (0)	4.0% (2)	18.0% (9)	78.0% (39)

This finding from interview data is backed up by data from the surveys. Survey data shows that 96% (n=48) of respondents rated the Talking Time© manual as being 'very helpful' or 'extremely helpful'. The plans and prompts for Story Conversations, Word Play, and Hexagons sessions were also well received, with 98.1% (n=53), 94.4% (n=51), and 96.0% (n=48) rating them as 'very helpful' or 'extremely helpful', respectively.

Interviews were insightful in highlighting what specifically was considered helpful about the resources connected to each activity. The use of wordless books for Story Conversations was very popular and repeatedly described as a novel and effective approach for engaging children in books, improving vocabulary and boosting confidence. In contrast, although the activity in itself was considered effective, the resources provided for the Hexagons activity were often criticised for poor quality and lack of relevance to children's everyday experiences and age group. For example, pictures that prompted conversations around getting ready for school were considered unrelatable for children in PVI settings in particular. Practitioners reported creating their own materials to ensure images were larger, higher quality, and more appropriate. School and maintained nursery settings in particular, praised the handbook's thoroughness. They felt that this made planning time-efficient and helped practitioners select a clear focus for each session.

Interviews with practitioners in PVI settings generated two distinct themes in relation to the perception and experience of resources: the level they felt resources were pitched at; and the closeness of resources to their existing practices. For some, the handbook was most helpful for junior colleagues who needed support on using a broader range of vocabulary with children, while one practitioner felt it was aimed at schoolteacher level, making it difficult to engage Level 3 practitioners in professional development. A trainer-mentor concurred with this sentiment, reporting that for their PVI settings the handbook was experienced as *'too much'*, inaccessible, and written for schools. It was considered too theoretical and the trainer-mentor believed a simple summary sheet reminded practitioners of core principles and ideas would have been sufficient. The second issue mentioned by two PVI settings was how closely resources aligned with their existing practices. For one setting, this meant that they rarely relied on the handbook to plan sessions, noting it felt like additional paperwork. Paying close attention to the handbook felt like an extra burden, and they considered themselves confident enough to adapt activities based on their own knowledge. The other setting viewed this alignment more positively, using the programme to become more intentional and structured in supporting their children's language development.

Trainer-mentors identified Story Conversations to be the most popular and successful activity, with its underlying principles and the utility of wordless books strongly embraced by practitioners. They observed practitioners adapting to this most easily, and the quick, visible impact on children's communication and confidence helped sustain practitioner enthusiasm. Many settings had plans to order additional wordless books to embed Story Conversations into their provision. Hexagons were also considered popular, with the picture prompts (when considered relevant) found to be effective in developing children's narrative skills. In contrast, Word Play was considered the most challenging activity for practitioners to understand and implement. Trainer-mentors attributed this to its similarities to activities practitioners were already familiar with, namely, guided play, making it harder to distinguish how Talking Time©'s approach differed and why it was impactful. In addition, Word Play required a shift to a child-led play model, which was a significant adjustment for practitioners accustomed to directing and leading play.

Experience of professional development

Table 43 below shows how survey respondents rated the professional development support received from trainer-mentors. In response to the question: 'How useful did you find the professional development support you received from your mentor (twilight training sessions, in-person mentoring, and online mentoring)?' over 98% (n=53) of respondents provided a rating of 4 ('useful/helpful') or 5 ('very useful/helpful') and over. Furthermore, in interviews, practitioners generally viewed all modes of professional development delivery as valuable. Each activity was described as offering distinct benefits.

Table 43: Practitioner perceptions of usefulness of professional development

	1 (Not at all useful) % (n)	2 (Somewhat unhelpful) % (n)	3 (Neither useful nor unhelpful) % (n)	4 (Somewhat useful) % (n)	5 (Very useful) % (n)
Professional development: How useful did you find the professional development support you received from your mentor	0% (0)	1.9% (1)	0% (0)	13.0% (7)	85.2% (46)

(twilight training sessions, in-person mentoring, and online mentoring)					

Twilight sessions were often described as useful and enjoyable by interviewees, particularly school practitioners who appreciated the opportunity to network with other early years professionals and ask questions in a supportive environment. Seeing videos of delivery was considered especially helpful, and the sequencing of twilight sessions before in-person mentoring was praised for providing an introduction to activities followed by opportunities for clarification. However, twilight sessions as a mode of professional development alone was considered less impactful than in-person or online mentoring for understanding how activities worked in specific contexts and were felt to be less interactive than other activities. Timing was also a concern, as sessions added to practitioners' workload, and occasional technical issues disrupted participation. Following twilight sessions, settings were given access to recordings of the online workshops to mitigate against any attendance issues, however, viewing these recordings at a later date was not mentioned by any settings interviewed.

In-person visits were consistently described as the most beneficial, with modelling of activities cited as integral to practitioners' understanding and confidence. Observing strategies in their own setting and receiving tailored guidance supported effective implementation, though absence of children at times limited the value of these sessions. Online mentoring, including video submissions, was also perceived as highly effective for professional development. While practitioners initially felt embarrassed recording and reviewing their practice, most reported that this became a powerful tool for reflection and improvement over time. Trainer-mentors were widely credited with creating a supportive, non-judgemental environment that encouraged openness and learning.

Trainer-mentors similarly highlighted that twilight sessions were the most challenging, as they took place after hours, often without additional pay, and raised concerns about how much information practitioners were able to absorb. The sequencing of activities was praised as well planned, allowing practitioners to move quickly from theory in twilight sessions to practice during in-person visits, where demonstrations helped address confusion and settle worries. Interestingly, while outlined above that some trainer-mentors considered more experienced practitioners to engage better with reflective practices one trainer-mentor noted that in-person mentoring was most challenging for experienced practitioners, who found it harder to identify and admit areas for development. These practitioners were more likely to view observation as a judgement of right or wrong rather than an opportunity for reflection. In contrast, online mentoring was described as a 'surprise hit', as the private space for reflection encouraged practitioners to open up about their practice and engage more deeply in self-assessment.

IPE RQ7: 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?)

Support from trainer-mentors (as outlined in IPE RQ6) and setting leads, and practitioner receptiveness to Talking Time© emerged in interviews as key contextual factors supporting the successful take-up of Talking Time© across settings. Additionally, successful team working was seen as important for enabling an equal distribution of programme learnings across Talking Time© leads and deep reflections on professional development.

Support from setting leads

Survey data shows a high rating of the support provided by setting leads in enabling and supporting the delivery of Talking Time©.

Table 44: Practitioners' ratings of support from senior leadership

	1	2	3	4	5

	(Not supportive at all) % (n)	% (n)	% (n)	% (n)	(Very supportive) % (n)
Supportiveness of senior leaders in enabling/supporting delivery of Talking Time©	0% (0)	3.8% (2)	7.6% (4)	26.4% (14)	62.3% (33)

Note: respondents were asked to rate support on a 5-point scale, ranging from "Not supportive at all" (1) to "Very supportive" (5).

Interestingly, the significance and patterns of support from senior leaders was captured best by interviews with trainer-mentors. Senior leadership support emerged as a critical contextual factor for the successful take-up and sustainability of Talking Time©. Trainer-mentors consistently highlighted that when leaders were actively engaged and understood the programme's demands and benefits, implementation was smoother, with practitioners allocated sufficient time to deliver sessions and participate in professional development activities.

In PVI settings, securing buy-in was reported as considerably easier due to flatter organisational structures, enabling leadership to quickly champion the programme and create a culture of enthusiasm around it. In contrast, gaining buy-in from senior leaders in schools was often more challenging because of longer chains of command. Trainer-mentors reported having to advocate for allocated time for Talking Time© practitioners to fully engage with the programme. A common observation was that early strategic conversations with senior leaders, especially in schools, were essential to ensure adequate time allocation for practitioners, integration into curriculum planning, and capacity building for future delivery.

Trainer-mentors observed that where leadership involvement was limited or superficial, such as signing up without fully appreciating the time commitment, practitioners were left to fit the programme around existing responsibilities, undermining fidelity (such as meeting the required number of sessions a week). Conversely, strong leadership engagement facilitated practical adjustments, such as pausing other interventions, and ensured that Talking Time© leads could transition into mentoring roles for subsequent cohorts, ensuring the sustainability of the programme.

A few interviews with setting leads did provide insight into their role during the programme. This insight generally came from school senior leaders who were more removed from day-to-day delivery than the PVI managers interviewed. While they expressed strong support for the programme, they did not feel the need to intervene or provide any additional support to that provided by trainer-mentors, which they regarded as excellent. One school lead did mention that they would have liked more time to reflect with the Talking Time© leads on their learnings and provide a structure for continuous reflection but scheduling constraints, particularly with support staff, meant this was not possible.

Practitioner receptiveness to the programme

As outlined above, practitioner openness to learnings from the programme is integral to it being taken up and implemented well.

Across settings, senior leaders reported that Talking Time© practitioners were open and enthusiastic about the programme, although some were initially concerned about the time commitment. These concerns were generally eased through effective timetabling. Interviews with practitioners across all setting types indicated that most felt positive and motivated to take part in Talking Time©, often linking their interest to concerns about low speech and language levels among children and high EAL intake. Many viewed the programme as a valuable way to address communication challenges and were particularly attracted to its use of wordless books and play-based activities.

Despite this enthusiasm, practical concerns were common, including fitting sessions into busy timetables, resource requirements, and staff capacity for planning and professional development. Some school practitioners described feeling initially overwhelmed by the size of the handbook and worried about the additional workload. However, most reported that these concerns eased with trainer-mentor support and time to plan and become familiar with the programme. This

highlights the importance of trainer-mentor guidance and dedicated planning time as key contextual factors in the successful take-up of Talking Time©.

Trainers-mentors generally reported strong receptiveness to the programme among practitioners across all settings. Where poor engagement was noted, such as lack of attendance at professional development sessions or failure to implement strategies, this was often linked to structural issues rather than practitioner attitudes. This included high staff absences and lack of senior leadership support, which affected capacity and enthusiasm. Interestingly, trainer-mentors observed that more confident and experienced practitioners were often harder to engage with Talking Time© approaches. Supporting these practitioners often required helping them with significant 'unlearning' to enable them shift towards a more child-led pedagogical approach, compared to less experienced colleagues who were typically more open.

Teamwork practices

Interviews with practitioners highlighted variations in team working practices across settings, with some notable differences between schools and PVIs. In schools, planning was often reported as being led by the Talking Time© lead who was the class teacher. They prepared the key vocabulary focus for the week, while leads who were teaching assistants were responsible for specific prompts and questions that would be suitable for the children in their group. Some school practitioners did report a more collaborative approach to planning Talking Time© delivery, meeting at the end of the week to review progress and plan ahead.

In PVIs, practitioners generally described a more integrated approach to team planning, with regular meetings to discuss resources and vocabulary for the week. However, barriers to effective teamwork were reported across setting types. Support staff in schools lacked planning, preparation, and assessment time and often worked limited hours, reducing opportunities for joint planning and formal reflection. In some cases, this led to difficulties delivering Talking Time© due to a lack of deep engagement with resources. Staff capacity limits across all settings further restricted collaboration in terms of session planning or formal reflection as covering continuous provision was a priority.

Despite the barriers identified in interviews, survey data does show consistent discussion and planning of activities among Talking Time© leads with 69% of respondents engaging in this weekly or daily. However, interview data provides insight into the structure of these discussions/planning sessions. As noted above, informal discussions are a common approach as opposed to formal planning sessions. Importantly, a significant number of practitioners report they are only able to take part in these discussions and/or planning sessions a few times a year. This reflects findings from interviews that finding time to do this is a challenge for many settings.

Table 45: Reported frequency of discussions/planning of Talking Time© activities among delivery leads

Frequency	% (n)
Daily	25.5% (14)
Weekly	43.6% (24)
A few times across the programme	29.1% (16)
Never	1.8% (1)

IPE RQ8: 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?

The only group of children that was referred to facing barriers to the programme were some children with significant SEND that impacted their ability to concentrate. Practitioners described initially trying to involve these children but that their inclusion was not sustainable as a 15-minute session was too long for them to be able to engage effectively. A PVI practitioner described an experience that represented this experience that was shared among several interviewees:

I checked before we started and said that we had a couple of SEND and was told, 'No, no, still add them in.' But they haven't engaged at all. Which is fine, I didn't actually expect them to, they're both non-verbal anyway. We had a few times that they'd maybe sit and actually look at the book fleetingly, which was a start, and I got all excited, and then that just didn't continue again. (TP06, PVI practitioner)

IPE RQ9: 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?

Based on the survey data, Story Conversations was considered the most engaging activity for children, with 67.9% (n=38) of respondents rating it as 'very engaging' (Table 46) and the most beneficial activity, with 61.7% (n=34) of respondents considering it so (Table 47). Hexagons was the second activity most likely to be considered 'very engaging' 50.9% (n=28), although closely followed by Word Play (45.5%, n=25) (Table 46), and 7.3% (n=4) and 5% (n=3) of participants considered these activities the most beneficial (Table 47). The last column in Table 46 has been referred in this case to draw out the nuances of ratings of different activities to compare.

Table 46: Practitioner perceptions of child engagement per session type

Session	Not at all engaging % (n)	Slightly engaging % (n)	Somewhat engaging % (n)	Mostly engaging % (n)	Very engaging % (n)
Story Conversations	0% (0)	0% (0)	0% (0)	32.1% (18)	67.9% (38)
Word Play	0% (0)	0% (0)	3.6% (2)	50.9% (28)	45.5% (25)
Hexagons	0% (0)	3.6% (2)	14.6% (8)	30.9% (17)	50.9% (28)

Table 47: Practitioner reported most beneficial activity

Beneficial activity	% (n)
Story Conversations	61.7% (34)
All equally	25.5% (14)
Hexagons	7.3% (4)
Word Play	5% (3)

In interviews, Story Conversations and Hexagons were reported as the activities that garnered the most positive engagement. Children were considered to flourish in small groups—and quiet children greatly valued the opportunities to be heard. When engagement was an issue, it was often attributed to children's unfamiliarity with the scenarios in certain books rather than their level of need. For example, Lulu Loves Flowers was frequently mentioned as challenging because activities like visiting a library or garden centre were unfamiliar to some children from disadvantaged backgrounds. To address this, practitioners created opportunities for children to experience these activities, such as library visits or implementing Word Play before Story Conversation to build context.

Interestingly, survey data shows more ambivalence towards Hexagons than any other activity with 18.2% (n=11) considering the activity 'somewhat engaging' or 'slightly engaging'. This echoed insights from interviews outlined earlier with Hexagons on the one hand considered an extremely useful approach to building narrative skills but on the other hand, the quality (i.e. size and picture quality) and suitability of pictured activities considered inadequate.

In interviews, Wordplay was noted as the most difficult activity for engagement, as it was very close to existing activities and children struggled to engage in play without adult direction as they were used to. The perception of Word Play in comparison

to other activities is reflected in the survey data with it being the activity least likely to be considered 'very engaging' (45.5%, n=25, compared to 67.9%, n=38 for Story Conversations, and 50.9%, n=28 for Hexagons; Table 46) and with the lowest number of practitioners considering it the most beneficial activity (5%, n=3) (Table 47). Despite this perception in comparison to other activities, it is worth noting that overall, 96% of practitioners considered this activity 'mostly engaging' or 'very engaging'.

Practitioners also reported challenges for some children with additional learning needs, who found it hard to sustain attention for a full 15-minute session. In these cases, sessions were shortened or broken into smaller parts. Interestingly, Word Play was reported as best at capturing the attention of children with SEND more effectively than Story Conversations, as it required less sustained focus and involved props and play elements. In some instances, non-verbal children's needs were too high for them to engage at all. Some PVI settings observed that engagement from EAL children grew over the course of the programme, with their English and confidence improving over time.

Perceived impact

IPE RQ10: 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?

Overall, taking part in Talking Time© was considered to have resulted in increased understanding, knowledge, and confidence by practitioners, setting leads, and trainer-mentors interviewed. This was reflected in both survey and interview data. Over 98% (n=54) of survey respondents reported a significant impact on their skills (rated 4 or above) (Table 48). In interviews, these gains were attributed both to participation in professional development activities and to delivering Talking Time© sessions.

Table 48: Practitioner perceived impact in increase in skills to support young children's oral language development

	1 (Not at all) % (n)	2 % (n)	3 % (n)	4 % (n)	5 (A great deal) % (n)
Impact on oral language development skills	0% (0)	0% (0)	1.8% (1)	49.1% (27)	49.1% (27)

Note: respondents were asked to rate perceived impact on a 5-point scale, ranging from "Not at all" (1) to "A great deal" (5).

In interviews, practitioners across setting type frequently highlighted a deeper appreciation of child-led practice, and the development of a skill set that allowed them to facilitate practice where children's thoughts and observations guided activities. This shift was seen as particularly valuable in engaging children with books and stories, with wordless books emerging as a key enabler.

> *We were both a little bit unnerved by wordless books at the beginning of this programme, but are now not fazed by them at all, and can really see the value in how we can use those.* (TP03, School-based practitioner)

> *I did my focus at university on early language and support around children's PSED [Personal, Social, and Emotional Development] within books. So, I've always loved stories, but I never thought to just not read it, to just look at the pictures.* (TP23, School-based practitioner)

In settings where practice was already closely aligned with business as usual, the programme was noted for strengthening practitioners' capacity and confidence to engage in critical reflection, both individually and collectively.

> *Although these are things that we were doing, it's made it more intentional, bringing it more to the forefront of our practice [...] We're more reflective on each other, more critical.* (TP01, PVI practitioner)

Setting leads observed improvements in confidence around storytelling and greater self-awareness in practice. One school lead described how a practitioner became more deliberate about embedding vocabulary during interactions with children:

> *Now, I feel like sometimes she's going into play thinking, 'Oh, I want to make sure I'm embedding these key words,' whereas before we would have them but they are just not as focused.* (SL09, School-based setting lead)

Trainer-mentors also reflected on changes in practitioners' self-perception, noting increased awareness of strategies to meet diverse children's needs and a stronger sense of themselves as learners. This mindset fostered engagement in professional development and allowed practitioners more time and ability to reflect on the intent and impact of their practice. They became more confident in delivering child-led sessions and self-correcting when needed.

> *I noticed that particularly one setting by the end, it was fantastic practice, really more relaxed, more child-led, more allowing children to come up with their ideas, and even if children were talking about something completely different, I could see the practitioner following that and then slowly, gently bringing children back into the main thread." (TM02)*

IPE RQ11: 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©?

Across most interviewed settings, practitioners, setting leads, and trainer-mentors have reported that practices introduced through Talking Time© have begun to cascade beyond the sessions themselves, influencing wider approaches to early years language and communication development. This indicates that one of the key aims of the programme—for activities and approaches taught by the programme to become embedded in general practice—is being realised in some settings.

In interviews with practitioners and trainer-mentors, a common activity reported as being adopted more broadly was Story Conversations. Many settings chose to embed child-led engagement with wordless books as a key activity and had ordered additional books to support this. For example, a practitioner described how in Story Conversations an activity now forms part of their setting's daily routine:

> *Our routine is they come in, they pick their name out and then they sit down on the carpet and they look at a story…when all of the children are in, we'll go and sit down and we go, 'Who would like to tell us about their story?'* (TP23, School-based practitioner)

This approach was evident across different contexts, with PVIs often embedding Story Conversations into practice with younger children and schools extending its use to older age groups. A school lead, who also teaches Reception, reflected on the impact on their own practice:

> *I've noticed an impact on my own practice even with the Reception children, that when we are sharing a story at the end of the day I'm not necessarily just reading the story. We are stopping, we're looking, we're asking questions, we're making links, we're using some of the strategies that way.* (SL07, School-based setting lead)

Embedding a child-led approach to language and communication was also reported as taking place in more subtle ways. Practitioners described being more patient, allowing children additional time to process and respond to prompts before intervening. Several also noted a greater tendency to engage in extended conversations with children, reflecting a shift towards more intentional and responsive interaction.

IPE RQ12: What 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 (e.g. EYPP eligibility) and /or EAL?

Perceived impact on children's language and communication skills

The vast majority of survey respondents (94.5%, n=52) perceived the impact of Talking Time© on children as 'very significant' or 'significant' (Table 49).

Table 49: Practitioners perceived impact of Talking Time©

	% (n)
Very significant	63.6% (35)
Significant	30.9% (17)
Moderate	5.5% (3)
Weak	0.0% (0)
Very weak	0.0% (0)

Similarly, the majority of practitioners interviewed believed that Talking Time© had a positive impact on children's language and communication. The distinction between different groups of children was commonly based on confidence and language skills rather than based on social characteristics (there was some commentary on children with EAL in interviews —see below). Practitioners reported that children who were already confident speakers, greatly improved their listening skills and ability to engage in interaction. For those with higher language and communication needs, their confidence to speak and communicate was perceived as significantly enhanced.

> *Usually, there's lots of bickering, and everybody just does their own thing really. This year we were amazed by how quickly they solved that problem, and how they could talk to each other, and explain what they needed to do, and listen to each other.* (TP17, School-based practitioner)

> *A lot of the children have a lot more confidence, which I think is half the battle with children this age— actually having the confidence and the willingness to have a go.* (TP23, School-based practitioner)

Practitioners described improvements as evident in several ways. Some noted progress in end-of-term assessments, but most highlighted changes observed in everyday communication. These included children using more sophisticated vocabulary, engaging in increased interaction and co-operation with peers, speaking with clearer articulation, and forming more complex sentences.

However, some practitioners cautioned against attributing all progress solely to Talking Time©, noting that improvements were difficult to separate from the development that would have occurred over the year even without the intervention. Two PVI respondents questioned the programme's impact outright. One reported no discernible impact, while another suggested that limited impact was likely due to the cohort having higher needs than the programme could fully address:

> *I just felt that the children we'd got this year weren't ready for it. A group, maybe three or four years ago, would have been, but more recently, the children coming in are more and more delayed in speech and communication.* (TP06, PVI practitioner)

Differences among groups of children

Survey data reflects insights from interviews above with regards to level of impact across different groups of children. Respondents felt that children who are quieter or less confident and with weaker language skills were those who benefited most from the intervention.

Table 50: Practitioners' perception of children who benefited most from the intervention

	% (n)
Children with weaker language skills	41.2% (21)
Children with EAL	11.3% (7)

Children with SEND	0% (0)
Children who are quieter or less confident	45.5% (25)
Not sure	2% (1)
Other	0% (0)

In interviews, when patterns were observed across social demographics, improvements in the ability and confidence of children with EAL to speak in English were consistently highlighted.

> *Our EAL children have made massive impacts from it [...] So, genuinely we had never heard their voices before and now the minute they sit down for sessions, they're chatting away like no-one's business.* (TP19, School-based practitioner)

> *I'm thinking of a little boy in particular who, his speech is coming, and he's saying much more in English [...] I don't think we ever would have heard that from him until maybe a lot later on.* (TP04, Maintained nursey school practitioner)

In terms of the least impact, a common observation from interviewees across all setting types was that children with the highest needs (often reported as children who were non-verbal and with pronounced difficulties in the ability to concentrate) were least affected by the programme, as they struggled to engage with the activities. Survey results reflect this with 11.3% (n=7) survey respondents categorising children with EAL as largest beneficiaries compared with no respondents considering children with SEND as such.

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

When asked to reflect on any unintended consequences or negative effects of Talking Time©, responses were overwhelmingly positive. Practitioners frequently noted improvements in children's behaviour, including greater patience, willingness to listen to one another, and enhanced attention and concentration. Mixed-ability grouping was described as surprisingly beneficial for lower-ability children, who gained confidence and skills by modelling higher-ability peers. Practitioners also reported that delivering sessions in this way gave them a deeper understanding of their children, enabling more accurate assessment of quieter children's abilities. Several highlighted that small-group conversations helped identify previously unrecognised speech and language needs, allowing them to place children on pathways for additional support.

Across interviews, practitioners and setting leads spoke about the unexpected and joyful outcome of children's growing love of books and independent engagement, often initially conversations with peers about stories. As one practitioner and setting lead reflected:

> *What's been really lovely is seeing children in provision together looking at books they've never seen before, and they'll be sitting together and pointing things out. I don't feel like I've ever seen children talk around books as much independently as I have before doing this.* (TP04, Maintained nursey practitioner)

> *They sit with their friends and read the book to their friends and say, 'What's happening?' Like, they mimic what we're saying and it's just nice to see.* (SL06, PVI setting lead)

In PVIs specifically, the programme was observed to have an indirect impact on two-year-olds, as the often open-plan nature of these settings exposed younger children to the activities and teaching styles used with older groups. Practitioners believed this meant they would enter the pre-school room at a higher level than previously expected in the next academic year.

One practitioner from a state-maintained setting did express surprise at the challenge posed by removing a practitioner from continuous provision to deliver Talking Time© sessions, namely having fewer staff to attend to children's needs and manage behaviour.

Usual practice

IPE RQ14: What was usual practice in relation to oral language development in all 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?

Usual practice in intervention settings/differences between setting types

The setting-level survey indicated that 57% of settings were implementing some oral language intervention during the same academic year. A slightly higher proportion was observed in the control group (62%) compared to the intervention group (52%). However, it is important to note that valid data were not obtained from all settings and that these percentages are based on responses from 84% of settings (103 out of 123).

Across settings that received the intervention, oral language development was already described as a priority in interviews and often cited as a key reason for wanting to take up Talking Time©, although approaches varied by setting type. In school-based and maintained nursery settings, most had pre-existing practices centred on conventional adult-led story reading for larger groups, while smaller group or individual reading was less common. These settings typically emphasised vocabulary reinforcement and book-led approaches, alongside creating 'language-rich' environments through frequent talk, modelling language, and encouraging repetition. School leads in particular noted challenges in supporting children who entered behind expected levels and in building the confidence of more junior staff to engage in extended discussion with children. Based on what was reported in interviews, PVIs also prioritised oral language development, commonly through storytelling and discussion practices that encouraged children to listen and respond. These settings more often noted that they incorporated child-led approaches in some oral language activities. Group sizes varied, with some settings favouring small groups delivered by key workers, although these were not explicitly described as mixed ability.

Many intervention settings had experience with language and communication programmes prior to Talking Time©, typically targeted at children with SEND or EAL. In schools and state-maintained settings, these included WellComm, Early Talk Boost, Concept Cat, and the Nuffield Early Language Intervention (NELI). One school used WellComm to screen all children due to high levels of developmental delay. PVIs generally reported fewer interventions, with one mentioning Lift Off while another referenced WellComm and an in-house speech and language therapist. One maintained nursery school practitioner recounted that their setting also once had an in-house speech and language therapist but found the process of removing children for one-on-one sessions unsettling and disruptive to their routines.

As well as programmes aimed at children, survey data revealed that it was somewhat common for intervention settings to have had previously received professional development related to supporting children's language development.

Table 51: Setting enrolment on to professional development activities other than Talking Time© to support children's language development

	% (n)
Yes, externally funded	8.9% (5)
Yes, internally funded	39.3% (22)
No, but planned before end of summer	0.0% (0)
No, but planned for next academic year	7.1% (4)
No	44.6% (25)

Comparison of Talking Time© to usual practice

Across interviews, practitioners generally described Talking Time© as accessible and effective compared to usual practice, though the nature of differences varied by setting type. In schools, practitioners noted that while storytelling was already

part of their routine, Talking Time© introduced a different approach. This included moving away from rigid story reading toward more conversational, child-led sessions, using smaller mixed-ability groups rather than whole classes, and focusing on discussing pictures after the story rather than using them to structure the narrative. Several practitioners felt this created a more natural atmosphere where children could follow their own ideas and choose groups, which they perceived as increasing engagement. As one practitioner reflected:

> *Although I felt like the skills that we've got were already really good, it was about approaching it in a different way and seeing whether or not that would give us some different responses*. (TP10, School-based practitioner)

In maintained nursery school settings, Talking Time© was described as using familiar resources like books but differing in its emphasis on conversation rather than simply reading through stories. Practitioners valued the small group, child-led format and noted that Talking Time© was aimed at all children rather than being an individual intervention, which they felt was less disruptive than approaches that removed individual children from play.

In PVIs, Talking Time© was seen as both similar and different to prior practice. While storytelling in groups was already common, practitioners highlighted that Talking Time© added structure and intentionality through activities such as the Hexagons task and mandated mixed-ability grouping. This was perceived as expanding practitioner tools and making their approach more deliberate. One practitioner explained:

> *These were things that we were doing, but it was unintentional, and now after the programme you're more intentional…you know which strategies to use for which child*. (TP01, PVI practitioner)

Setting leads generally agreed that Talking Time© was not difficult to adopt because it aligned with existing practices, and some noted that the books mapped well onto their curriculum. One school setting lead commented that this universal design made Talking Time© easier to implement than targeted interventions. Several PVIs highlighted that Talking Time© provided clear strategies and terminology, making it easy to follow, though it required additional administrative work than other interventions. Much of the administrative work referred to was in relation to the research project, such as uploading registers, rather than the intervention itself.

From the trainer-mentors' perspective, they felt that Talking Time© was unique compared to any professional development they had delivered as it combined core principles with flexibility, which offered scaffolding and support while allowing adaptations to suit different contexts. They felt this structure, combined with a non-judgemental approach, made the programme sustainable.

Usual practice in control settings/changes in practices during the trial

Several control settings interviewed reported prioritising language and communication development within their early years curriculum, often due to contextual factors such as high proportions of children with EAL and the additional needs associated with socio-economic deprivation. There was considerable variation in the approaches described, which included using the ShREC (Share, Respond, Expand, Conversation) approach, delivering small-group sessions focused on language development, incorporating singing as a tool to build confidence and language skills, and centring quality interactions in their practice. For children with EAL specifically, one setting explained an approach of starting with basic vocabulary at the single-word level and gradually building complexity. Control practitioners also described strategies such as adopting an emotional literacy perspective to support children to develop strategies to express themselves and using storytelling to enhance communication and vocabulary, with one practitioner noting, this included books without words in their setting.

Most control settings interviewed had previously implemented or were currently using language and communication interventions, typically targeted at children with SEND or EAL. These included programmes such as Language Box for children with no communication, Words First for one-on-one interventions with SEND children, Talk With Tales for Children

(TWiTCH) for professional development and targeted support, WellComm, ECAT, Colourful Semantics, and EAL-specific interventions.

Although not selected for Talking Time©, in interviews, two control settings reported making changes to their practice as a result of interaction with the programme. One control setting lead described how having watched the assessors come in, to do initial assessments with children, their strategy of giving children time to respond to conversation prompts had been taken on by them and their team. In the interview, they interpreted this strategy from Talking Time© experts. Following this visit, they had begun advising their practitioners to make a concerted effort to give children more time to answer questions. Another control setting lead spoke about actively seeking other professional development opportunities without access to Talking Time©. Interestingly, one control setting lead, noted that they made a deliberate effort to avoid any changes in order to remain true to their assignment in the control group.

Cost

We report in this section the cost of Talking Time© from the perspective of the implementing setting. Social costs, such as any additional time spent by parents/carers in supporting the programme, are not included. However, no such costs were envisaged at the design stage or reported during qualitative interviews. Costs incurred by schools are included as staff time required for training. No additional costs were investigated, although school staff provided general assessment of the demands of the programme, which are discussed later in this section.

Talking Time© is essentially a professional development programme, whose main components are described in Table 52. Staff from the delivery teams train mentors who, in turn, train setting staff to deliver the activities. The ingredients of the interventions are detailed in Table 53 for an average setting with 18 children.[11]

Where time inputs are reported in days, one day is assumed to be equivalent to 7.5 hours of work.

Table 52: Talking Time© professional development activities

Talking Time© Professional Development		
Introductory sessions two x one hour	Online briefing on oral language development One welcome session (face-to-face in the setting)	October 2024 – November 2024
3 x 1.5 hours twilight workshops *(all staff delivering Talking Time© sessions)*	Online training to introduce staff to the three Talking Time© activities, the manual, and materials, underpinning language-supporting strategies and reflective framework for ongoing professional development	December 2024 – March 2025
Four x one-hour in-class mentoring *(Talking Time® leads i.e. practitioners leading on delivery)*	These practical, supportive sessions build confidence and expertise in leading Talking Time© sessions and using tailored language-supporting interactions to enhance each child's oracy skills The mentor and practitioner take turns leading a session and reflecting on them. Later sessions support planning and adaptation for sustainment	Three sessions December 2024 – March 2025 One further session May 2025 – June 2025
Three x one-hour video mentoring *(the team's pedagogical lead for Talking Time®)*	These online sessions support ongoing leadership, planning, and adaptation of the programme longer-term. Some involve shared reflections on Talking Time© sessions video-recorded by the participating staff member	January 2025 – February 2025 February 2025 –March 2025 April 2025 – May 2025
Weekly individual reflection *(setting practitioners)*	In addition to the professional development support provided, practitioners are asked to engage in a weekly individual professional reflection on their language enhancing practices. A framework is provided to scaffold this reflection and practitioners select their preferred format, e.g. collaborative discussion, individual notetaking	Throughout

Table 53: Ingredients of Talking Time©

Category	Ingredient
Personnel for preparation and delivery	**Setting staff time for delivering the Talking Time© sessions**: Ten hours per group of five children, equivalent to 40 hours per setting for an average setting of 18 children. This reflects to two 15-minute sessions per week for 20 weeks for each group of five children
	Delivery team time for delivery supervision: 1.5 days per setting over the programme

[11] Note that there is no statutory class size for nurseries and that there are no official statistics on the average class size. We use our calculations for the average class size observed in the settings included in the study. Note that this figure has no bearing on the calculation of project costs (classroom delivery cost is set to zero because the intervention is embedded in the setting practice).

Personnel for training	**Setting staff time for training**: 12.5 hours per setting for training two setting leads comprising one-hour welcome visit, three 1.5-hours of twilight training sessions, four one-hour in-class mentoring sessions, and three one-hour online mentoring sessions
	Setting staff time for reflection: 7.5 hours per setting for two setting leads covering weekly planning, adaptation of plans, reflection, and peer collaboration
	Delivery team time for mentoring: 4.5 days of delivery team mentor time per setting to deliver the 12.5 hours of setting lead training, including preparation and follow-up
	Delivery team time for mentor administration and contingency: One day of a delivery team mentor per setting for general administration, contingency (rescheduling, troubleshooting), and additional support
Training costs	**Delivery team time for training mentors**: 1.1 days of delivery team staff per setting (66.5 days in total) to train mentors
	Travel/subsistence for mentors: In the train-the-trainer phase
Programme costs	**Delivery team time for programme co-ordination and management:** Two days of delivery team staff per setting (128 days in total)
	Delivery team time for support and supervision of mentors: 0.5 days of per setting (29 days in total)
Facilities, equipment, and material	**Materials for settings:** Practitioner handbook, Hexagons resources pack, story books, and welcome pack
	Materials for mentors: Handbook, programme resources, and World Play material
Other programme inputs	**Postage:** Of welcome packs, handbooks, and materials to settings ahead of programme
	Travel/subsistence for mentors: For welcome visits and in-person mentoring visits

The costs of the programme primarily consist of training costs. These include delivery team time spent delivering the training, setting staff time spent receiving training, and travel and associated travel and subsistence expenses. Delivery of the intervention itself does not generate additional costs because it is embedded within standard practice and delivered during usual teaching hours. Additional costs were incurred by the delivery team for supervising implementation and providing ongoing support. Supporting material were also distributed to settings and trainers to facilitate group activities.

Table 54 presents the costs per setting of each ingredient over a three-year period. For each year we report the quantity of each ingredient, its unit cost, and the resulting cost per setting. We distinguish between start-up costs, which are incurred only in the first year, and recurring costs, which are incurred annually.

The following costs are set to zero, because these will not be incurred by schools. Classroom delivery is costed at zero because Talking Time© will be embedded within nursery practice and will not require additional staff time. Mentor training is also assumed to occur at no cost to schools. The costs associated with training mentors (including travel and subsistence) were borne by the delivery team for the purpose of the trial and included the development of the training programme. We assume that these costs will be substantially lower in future implementation and are, at present, captured under the item 'Talking Time© support, supervision, and ongoing mentor training'.

Training costs are not confined to the first year, but they do not occur at the same level in subsequent years. We assume that staff turnover and refresher training will require additional training equivalent to 25% of the initial training input in years two and three of the programme. These costs are assumed to be fully absorbed by the settings and to involve no external facilitators.

Although the project budgeted £ 7.50 per hour as a contribution to setting staff time, we used a more realistic rate of £15.00 per hour, reflecting the cost a setting would incur if implementing the programme independently. We further assume that training materials can be reused over the three-year period, and do not require replacement.

Table 54: Cost of implementation of Talking Time© by ingredient

Category	Ingredient	Type	Year one			Year two			Year three		
			Mean quantity	Price per unit	Mean cost	Mean quantity	Price per unit	Mean cost	Mean quantity	Price per unit	Mean cost
Personnel for preparation and delivery	Classroom delivery	Recurring	40 hrs	0	0	40 hrs	0	0	50 hrs	0	0
	Supervision	Start-up	0.21 days	£400	£84						
Personnel for training	Staff cover	Recurrent	40 hrs	£15.00	£600	10 hrs	£15.00	£150	10 hrs	£15.00	£150
	Talking Time© leads training	Start-up	4.5 days	£400	£1,800						
	Talking Time© leads admin	Start-up	0.5 day	£400	£200						
	Travel and subsistence leads training	Start-up	1	£77.07	£89.17						
Training costs	Talking Time© mentor training	Start-up	1.07 day	£0	£0						
	Travel and subsistence mentor training	Start-up	1	£0	£0						
Programme costs	Talking Time© co-ordination and management	Recurring	1 day	£315.55	£315.55	1 day	£315.55	£315.55	1 day	£315.55	£315.55
	Talking Time© support, supervision, and ongoing mentor training	Recurring	0.47 day	£400	£187.10	0.47 day	£400	£187.10	0.47 day	£400	£187.10
Facilities, equipment, and material	Package of material for settings	Start-up	1	£194.95	£194.95						
	Package of material for mentors	Start-up	1	£12.90	£12.90						
Other inputs	Postage of material	Start-up	1	£21.85	£21.85						

Total cost of implementing Talking Time© per setting for the first year was estimated at £3,505.52 and projected at £4,810.82 over a period of three years (see Table 55 and Table 56). This corresponds to a cost per child per year of £64.25 over three years.

The three-year projections include a recurring cost for training new staff and retraining existing staff, set at 25% of the cost incurred in the first year. The programme does not currently envisage covering these costs. However, given the high levels of staff turnover in nurseries, it is difficult to see how such costs could be avoided unless the programme were to accept lower quality delivery in years two and three, which would be likely to affect outcomes. We therefore, assume that these costs will be absorbed by the settings.

It is not clear however that all settings will be able to absorb these costs. Our survey of setting leads indicates that implementation of Talking Time© was perceived as 'easy' or 'very easy' in 21.8% settings, 'manageable' in 54.6%, and 'difficult' or 'very difficult' in 23.6% of cases. When asked about the reasons for implementation difficulties, setting leads consistently cited staff shortages and the additional time required by the programme in addition to existing practice. Although our cost calculations make no explicit allowance for additional setting staff time, these findings suggest that staff capacity could be a constraint to programme implementation in years two and three for a substantial proportion of settings.

Table 55: Costs of Talking Time©

Item	Type of cost	Cost	Total cost over three years	Total cost per child per year over three years
Classroom delivery	Recurring	£0	£0	£0.00
Supervision	Start-up	£84	£84	£1.12
Staff cover	Recurrent	£600	£900	£12.00
Talking Time© leads training	Start-up	£1,800	£1,800	£24.00
Talking Time© leads admin	Start-up	£200	£200	£2.67
Travel and subsistence leads training	Start-up	£89.17	£89.17	£1.19
Talking Time© mentor training	Start-up	£0	£0	£0
Travel and subsistence mentor training	Start-up	£0	£0	£0
Talking Time© co-ordination and management	Recurring	£315.55	£946.65	£13.00
Talking Time© support, supervision, and ongoing mentor training	Recurring	£187.1	£561.3	£7.48
Package of material for settings	Start-up	£194.95	£194.95	£2.60
Package of material for mentors	Start-up	£12.9	£12.9	£0.17
Postage of material	Start-up	£21.85	£21.85	£0.29
Total		**£3,505.52**	**£4,810.82**	**£64.25**

Table 56: Cumulative cost per setting (assuming delivery over three years)

	Year one	Year two	Year three
Talking Time©	£3,505.52	£4,162.17	£4,810.82

Conclusion

Table 57: Key conclusions

Key conclusions	
1.	Children in Talking Time© settings made one month's additional progress in language development, on average, compared with children in other settings. This result has a high security rating.
2.	Among children eligible for EYPP, those in Talking Time© settings made approximately two additional months' progress in language development compared to those in other settings. Similarly, children with EAL made around two extra months' progress in English language development relative to those in other settings. However, these results may have lower security than the overall findings because of the smaller number of children analysed.
3.	Among the different components of language development assessed, Talking Time© appeared to have the strongest impact on expressive vocabulary. There was less evidence of an impact on grammar overall, although subgroup findings suggest the programme may have been particularly beneficial for grammar skills among children with EAL and children eligible for EYPP.
4.	Implementation of Talking Time© largely occurred as planned. The programme was well attended by children (80% of children attended at least 32 sessions—80% of the programme), and variation in attendance did not affect average programme impact.
5.	Talking Time© is designed to be embedded in standard nursery practice and sustained independently. Practitioners reported benefits from the professional development, perceived improvements in children's oral language, and confidence in continuing delivery. However, challenges around staff capacity, particularly turnover, and limited flexibility in timetabling and session content suggest further adaptation and training may be needed to sustain delivery in years two and three.

Impact evaluation and IPE integration

Implementation fidelity was high. Practitioners reported being able to adhere to the planned delivery model, with limited adjustments, in terms of sessions frequency, timetabling, and grouping of children by ability. Child attendance as reported in the data logs provided by the settings was very high. Taken together, these findings suggest that any absence of overall programme impact cannot be attributed to weak implementation or low levels of participation.

Despite the high levels of attendance observed in the data, practitioners reported that child absence presented a practical challenge, particularly in relation to delivering two sessions per week. High attendance was often achieved through rescheduling and catch-up sessions, indicating that a more flexible timetable may better align with the needs of nursery settings.

Talking Time© is an integrated programme comprising flexible activities for children and professional development aimed at improving staff understanding, confidence, and practice in supporting oral language. Practitioners, setting leads, and mentors consistently reported improvements in practitioner understanding, confidence, and skills in supporting early language development. Changes in practices extended beyond formal sessions: Changes in practices extended beyond formal sessions and some elements became embedded with everyday practice

The success of the professional development programme was particularly evident in the in-person and online mentoring components, which were highly valued by practitioners during interviews. In contrast, written reflections were used less frequently, and the programme was relatively less effective in promoting self-reflection. The primary reasons for this appear to be time pressures and competing demands on staff.

The programme had a positive impact on an index capturing multiple dimensions of language development, reflecting the range of activities delivered (including Word Play, Story Conversations, and Hexagons sessions). The improvement in the index was largely driven by gains in expressive vocabulary, which provides a relatively conservative measure of children's language skills. Many studies focus on receptive vocabulary (the ability to understand words) rather than expressive vocabulary (the ability to produce words), and it is not uncommon for studies to assess children's recognition of words that were explicitly targeted by the intervention. In this context, expressive vocabulary constitutes a more stringent test of the effectiveness of the intervention.

The impact of the programme was not consistent across all the assessments and, in particular, we observed no impact on the GAPS test assessment. One possible explanation is that the observed impacts reflect the success of the different Talking Time© activities. The largest and most robust impact of the programme was observed on expressive vocabulary. This finding seems to be consistent with practitioners' reports of particularly strong implementation, and children's engagement, with the Story Conversations sessions, which are more likely to support this aspect of language development. A key challenge for the programme going forward is to support equally strong adoption of the other programme components (specifically the Hexagons and Word Play activities), which target additional language skills. Another possible explanation for the inconsistent impact of the programme across assessments is the concurrent implementation of other language development programmes in both intervention and control settings. While Talking Time© was described by practitioners as original and accessible, many settings reported implementing other language development programmes, often targeted at children with SEND and EAL. Our setting-level survey, although covered only 84% of setting, found that 57% were implementing oral language interventions other than Talking Time© during the academic year, with a proportion slightly larger in the intervention group than in the control group (62% vs 52%). This has some implications for the interpretation of the findings as an efficacy trial. In principle, efficacy trials seek to isolate the effects of the intervention under study. In this context, an absence of impact should not be interpreted as evidence that Talking Time© is not effective, because its impact could have been diluted by the presence of concurrent overlapping interventions.

Interestingly, the results for children with EAL and those eligible for EYPP show a different pattern. For these subgroups, the largest effects were observed on the grammar component of RAPT. In the absence of further evidence, this finding is difficult to interpret and should be treated with caution, given the small sample sizes for these groups.

Variations in usual practice between setting types may have influenced the outcomes. School-based and maintained nursey settings typically relied on adult-led, large-group story reading, whereas PVIs were more accustomed to small-group activities and encouraging child engagement during reading. These different starting points and levels of familiarity with child-led approaches may have contributed to differences in impact across setting types. We conducted an exploratory analysis comparing PVIs and non-PVIs, which found no impact of the intervention in PVI settings. This appears to support the interpretation that many of the of the practices promoted by the intervention were already in place in PVI settings.

The absence of impact in PVI settings is also consistent with an alternative explanation centred on differences in the characteristics of staff in these settings. The delivery team reported that professional development was more challenging to implement in PVIs, which may be less accustomed to reflective practice. In addition, variability in the attendance of both children and staff may have further reduced the effectiveness of the intervention in these settings. However, neither the quantitative data, nor the qualitative interviews provide sufficient evidence to fully explain the lack of impact in PVIs. It should also be noted that PVIs made up less than 20% of the sample, meaning the study was unpowered to detect small effects in these settings, and that there is considerable uncertainty around the estimated effects of Talking Time© in PVIs.

There is also some evidence of behavioural change in control settings potentially induced by the intervention. One setting reported modifying their practice around conversations following observations of children's baseline assessments and another actively sought out professional development opportunities after being assigned to the control group. Although this evidence is purely anecdotal it may indicate unobserved changes in behaviours within control settings that could bias the study comparison.

The intervention had a larger estimated impact among children with EAL and those eligible for EYPP. However, the sample sizes for these subgroups were small, and the results should therefore, be interpreted with caution. These findings are broadly consistent with practitioners' observations that the programme was particularly beneficial for children with language difficulties, though not necessarily for children with EAL specifically. For other groups, some practitioners reported difficulties in engaging children with more complex needs, which is not unexpected given that the programme is not currently designed to address more severe or complex language difficulties. Children with EAL and EYPP children are among those most at risk of delayed English language development and therefore, more likely to benefit from the intervention. Extending the programme to better support children with complex needs would likely require a redefinition of its activities.

The IPE identified several contextual factors that may moderate programme success, including the availability in the settings of quiet space, staff capacity, setting financial resources, and leadership support. These resources are likely to vary substantially across settings. However, the impact evaluation did not collect sufficient data to examine how these factors influenced programme impact in the quantitative analysis.

Practitioners expressed strong enthusiasm for Talking Time© and a willingness to continue implementation in future years, however, they also highlighted important challenges, consistent with the findings of our cost analysis. In particular, practitioners reported pressure on staff time in settings already operating at or near full capacity, as well as high levels of staff turnover, which limit the ability to retain trained staff and create a recurring need for retraining new staff.

These findings raise some questions regarding the sustainability of the programme delivery beyond the first year without additional resources. Addressing these challenges will be important for the delivery team. A planned follow-up study by the evaluation team in early 2026 in the same settings will examine whether, and under what conditions, these barriers can be overcome to support long-term sustainability.

This evaluation also offers some insights on how the intervention could be refined or improved in the future to increase its potential impact. The following recommendations emerge:

- **Introduce greater flexibility in session timing.** Children's attendance and staff shortages made delivery more challenging. A more flexible approach to timetabling sessions, both across the year and within each week, may improve attendance and support delivery.
- **Review the suitability of the intervention in PVIs.** Child-led activities and storytelling already appear to be common practice in PVI settings, suggesting that the intervention might be partially redundant in these settings. In addition, challenges in delivering professional development in PVIs, together with more variable attendance among both children and staff, may help explain this pattern.
- **Consider more targeted delivery at the setting level.** Focusing the intervention on settings with a higher proportion of children with language difficulties may increase its effectiveness. The programme was perceived as particularly beneficial for children with EAL and those eligible for EYPP, suggesting that targeting settings with larger numbers of children would maximise its population-level impact.
- **Reconsider the structure of the sessions delivered.** Joint storytelling appears to be the most successful component in terms of children's engagement and outcomes. The programme may benefit from rebalancing the emphasis placed on its components, particularly the relative weight given to Word Play and Hexagons activities.
- **Provide ongoing training and support to staff.** Given high staff turnover in nursery settings, the stock of knowledge gained through the initial professional development is likely to diminish over time. Offering additional training or refresher sessions after the first year could help sustain the quality of delivery. This could include providing recordings of training sessions (already available) to support the induction of new staff. In addition, a more structured approach would involve training all nursery staff, rather than only two programme leads, to deliver the intervention.

A follow-up qualitative study conducted between February and April 2026, several months after the trial had ended, provided further insights on sustainability beyond the trial period. Of the nine contacted settings, seven reported that they had continued implementation, while two stated that they did not intend to continue delivery. This pattern is broadly consistent with staff survey data, where 23% of respondents reported that implementation had been 'very difficult', and 11% indicated they may not continue. These findings suggest that while implementation was sustained in the majority of settings, a minority experienced challenges that affected continuation. Given the small sample size and the short period of time since the end of the trial, these results should be interpreted as indicative rather than conclusive, and longer-term follow-up would be needed to establish whether continuation reflects a temporary or more persistent pattern..

The identified challenges to continuation included: i) high staff turnover, the primary barrier to continuing the intervention is the loss of trained personnel; and ii) the flexibility of the timetable and programme content, where nurseries reported struggling to fit the intervention in their schedules. Key recommendations therefore included: developing a training manual or toolkit for staff unable to receive trainer-mentor support; introducing formal check-ins or light-touch support at the end of the programme to support quality assurance during independent implementation; extending Talking Time© training beyond the two leads to maintain knowledge within settings; and building a resource bank of books and other materials that could be accessed by staff for delivery.

Other findings from the follow-up study, however, point to encouraging longer-term effects. Staff reported an increase in reflective practice, which was not the norm during the trial, and indicated that this had become embedded in routine practice. There was also evidence that staff were incorporating aspects of the intervention, such as child-led group conversations, into everyday practice, even if not in a structured way, suggesting sustained change in practice. In addition, staff reported adapting intervention activities to fit setting schedules and children's needs. This flexibility further indicates that the intervention may be sustainable. The intervention was not designed to be rigidly structured; rather, it encouraged reflective practice with the aim of supporting adaptation to local conditions.

Limitations and lessons learned

Some floor effects were observed in the baseline assessments, particularly for the GAPS test measure, for which nearly 40% of children scored zero at baseline. Analysis of the baseline data indicates that this was largely attributable to the young age of the children in the study (who were just above the minimum eligibility threshold for the GAPS test) and to the introduction of a stopping rule during test administration. These floor effects did not, however, affect the estimated impacts. Robustness checks using alternative model specifications show that the absence of an impact on the GAPS test outcome is stable across approaches that account for missingness and censoring.

The decision to change the primary outcome from the GAPS test to a composite index, made after the preliminary analysis of baseline data, proved to be beneficial, because the index is better suited to capture changes across the different dimensions of language development targeted by the intervention. The use of the index resulting in only a limited loss of data, despite requiring complete information across three separate assessments. Specifically, 5.2% of children assessed at follow-up were not included in the index compared to 4.0% for the REV assessment and 1.7% for the RAPT assessment individually considered. This rate of missingness is also lower that that observed for the GAPS test assessment (9.0%).

The study included children eligible for EYPP and children with EAL at levels broadly consistent with their representation in the general population despite considerable efforts at the recruitment and randomisation stage to include children more likely to be disadvantaged. The study was not statistically powered to identify subgroup effects, and findings relating to subgroups should therefore, be interpreted with caution.

The study did not include a quantitative measure of improvements in practitioners' capacity to deliver Talking Time© sessions as outlined in the theory of change in terms of confidence, knowledge, skills, and language-supporting practices. An endline survey of practitioners to measure such capacity produced incomplete and limited observations. Progress in this area was instead explored through the process evaluation using qualitative interviews. The inclusion of a quantitative analysis of the impact of Talking Time© on intermediary outcomes of the intervention (practitioners' knowledge, confidence, and skills) could have helped interpreting the results and provided additional insights into an important component of the logic model. A study by the delivery team jointly funded by the EEF and the Nuffield Foundation based on audio-recordings of Talking Time© sessions, will shed some light on the intervention mechanisms.

The study assessed the impact of the intervention across different levels of compliance measured by attendance at Talking Time© sessions (held twice a week). However, one of the programme's objectives was to influence broader practice in supporting oral language development, and qualitative interviews suggest that practitioners did embed Talking Time© approaches in their routine practice. This implies that the impact of the intervention may be related to overall engagement with the programme as measured by overall attendance, rather than attendance of Talking Time© sessions alone. Unfortunately, the evaluation team did not collect data on overall attendance, and we were therefore, unable to investigate this further.

Similarly, we were unable to examine heterogeneity of impacts across settings. The IPE identified several potential moderators of impact, including the availability of space, staff time and capacity, financial resources, and leadership support. These factors are likely to vary substantially across settings. Settings varied in organisational structure (e.g. maintained vs PVI provision) and baseline characteristics, such as staff qualifications and turnover. Although setting-level information was collected to support this analysis, survey completion was incomplete, and the sample of complete responses was too small to support reliable statistical comparisons between settings with different characteristics. It should also be noted that the information collected from setting was based on self-report rather than direct observation and is therefore, subject to measurement error. Evidence of such error is apparent in cases where the same setting was reported on by different staff members, with responses differing across respondents.

Future research and publications

The current findings are broadly consistent with previous evaluations of Talking Time©. The intervention produces mixed effects across different indicators of language development, and the overall impact of one month's progress in oral language is consistent with impacts observed in previous evaluations of the same intervention. The inclusion of other assessments of language development, as well as indicators capturing other potential benefits suggested by the IPE, such as improvements in the quality of group interactions, could further our understanding of the overall effectiveness of the intervention.

The follow-up study identified some longer-term effects of the intervention, with elements becoming embedded in current practice. However, it also reported discontinuation in some settings, with staff turnover identified as the main barrier to continued implementation. A key question for future research is the extent and significance of these long-lasting reported changes in practice, and whether sustained implementation can be supported through light-touch ongoing support to the training of new staff.

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'*. London: Lawrence Erlbaum Associates, Inc. Available at: https://www.cal.org/wp-content/uploads/2024/04/cal_documents_executive-summary-developing-literacy-in-second-language-learners-report-of-the-national-literacy-panel-on-language-minority-children-and-youth_august-shanahan_2006.pdf (accessed 05/02/2024).

Data Protection Act 2018, c.12. Available at: www.legislation.gov.uk/ukpga/2018/12/contents (accessed 24/06/2026).

Dobinson, K.L., Mathers, S. and Dockrell, J.E. (2022) *'The Quality of Universal Language Support in Maintained Nursery Classrooms in Two Socially Disadvantaged Areas of England'*. Unpublished manuscript.

Dockrell, J.E., Howell, P., Leung, D. and Fugard, A.J.B. (2017) 'Children with Speech Language and Communication Needs in England: Challenges for Practice'. *Frontiers in Education*, 2, 35. https://doi.org/10.3389/feduc.2017.00035

Dockrell, J.E., Law, J., Mathers, S., Forrest, C., Charlton, J., Dobinson, K. and Hewitt, L. (2023) '*Empowering Staff to Enhance Oral Language in the Early Years: Cluster Randomised Trial: Final Report'*. London: UCL Institute of Education. Available at: https://discovery.ucl.ac.uk/id/eprint/10170459/ (accessed 31/01/2024).

Dockrell, J.E., Stuart, M. and King, D. (2010) 'Supporting Early Oral Language Skills for English Language Learners in Inner City Preschool Provision'. *British Journal of Educational Psychology*, 80: 4, 497–515. https://doi.org/10.1348/000709910X493080

Durham, R., Farkas, G. Hammer, C., Tomblin J., and H. Catts (2007) 'Kindergarten oral language skill: A 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 (EEF). (2022) *'Statistical Analysis Guidance for EEF Evaluations'*. London: Education Endowment Foundation. Available at: https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluation/evaluation-guidance-and-resources/evaluation-design (accessed 20/02/2024).

Education Endowment Foundation (EEF). (2023) *'Cost Evaluation Guidance for EEF Evaluations'*. London: Education Endowment Foundation. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/evaluation-design/Cost-Evaluation-Guidance-Feb_2023.pdf?v=1743094546 (accessed 23/02/2024).

Efron, B. and Hastie, T. (2016) *'Computer Age Statistical Inference: Algorithms, Evidence, and Data Science'*. Cambridge: Cambridge University Press. https://doi.org/10.1017/cbo9781316576533

Eldridge, S.M., Ashby, D. and Kerry, S. (2006) 'Sample Size for Cluster Randomized Trials: Effect of Coefficient of Variation of Cluster Size and Analysis Method'. *International Journal of Epidemiology*, 35: 5, 1292–1300. https://doi.org/10.1093/ije/dyl129

Gardner, H., Froud, K., McClelland, A. and Van Der Lely, H.K.J. (2006) 'Development of the Grammar and Phonology Screening (GAPS) Test to Assess Key Markers of Specific Language and Literacy Difficulties in Young Children'. *International Journal of Language and Communication Disorders*, 41: 5, 513–540. https://doi.org/10.1080/13682820500442644

General Data Protection Regulation (GDPR). (2016) *'Council Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data (United Kingdom General Data Protection Regulation) (Text with EEA relevance)'*. Available at: www.legislation.gov.uk/eur/2016/679 (accessed 24/06/2026).

Goldsmith, L.J. (2021) 'Using Framework Analysis in Applied Qualitative Research'. *The Qualitative Report*, 26: 6, 2061–2076. https://doi.org/10.46743/2160-3715/2021.5011

Golinkoff, R.M., Hoff, E., Rowe, M.L., Tamis-LeMonda, C.S. and Hirsh-Pasek, K. (2019) 'Language Matters: Denying the Existence of the 30-Million-Word Gap Has Serious Consequences'. *Child Development,* 90: 3, 985–992. https://doi.org/10.1111/cdev.13128

González, S. and Bonal, X. (2021) 'COVID-19 School Closures and Cumulative Disadvantage: Assessing the Learning Gap in Formal, Informal and Non-Formal Education'. *European Journal of Education*, 56: 4, 607–622. https://doi.org/10.1111/ejed.12476

Hassinger-Das, B., Toub, T.S., Hirsh-Pasek, K. and Golinkoff, R.M. (2017) 'A Matter of Principle: Applying Language Science to the Classroom and Beyond'. *Translational Issues in Psychological Science*, 3: 1, 5–18. https://doi.org/10.1037/tps0000085

Hayes, R.J. and Moulton, L.H. (2017) *'Cluster Randomised Trials'*. New York, NY: CRC Press. https://doi.org/10.4324/9781315370286

Hoff, E. (2014) 'Causes and Consequences of SES-Related Differences in Parent-to-Child Speech'. In: M.H. Bornstein, R.H. Bradley (Eds.), *'Socioeconomic Status, Parenting, and Child Development (pp. 1–304)'*. New York, NY: Routledge. https://doi.org/10.4324/9781410607027

Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J. and Hedges, L.V. (2010) 'Sources of Variability in Children's Language Growth'. *Cognitive Psychology*, 61: 4, 343–365. https://doi.org/10.1016/j.cogpsych.2010.08.002

I CAN. 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: I CAN and RCSLT. Available at: http://www.bercow10yearson.com/ (accessed 20/06/2024).

Imbens, G.W. and Rubin, D.B. (2015) '*Causal Inference: For Statistics, Social, and Biomedical Sciences: An Introduction'*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139025751

Kieffer, M.J. (2008) 'Catching Up or Falling Behind? Initial English Proficiency, Concentrated Poverty, and the Reading Growth of Language Minority Learners in the United States'. *Journal of Educational Psychology*, 100: 4, 851–868. https://doi.org/10.1037/0022-0663.100.4.851

Klingberg, S., Stalmeijer, R.E. and Varpio, L. (2023) 'Using Framework Analysis Methods for Qualitative Research: AMEE Guide No. 164'. *Medical Teacher*, 46: 5, 603–610. https://doi.org/10.1080/0142159X.2023.2259073

Law, J., Charlton, J. and Asmussen, K. (2017) *'Language as a Child Wellbeing Indicator'*. London: Early Intervention Foundation. Available at: https://foundations.org.uk/wp-content/uploads/2026/03/language-child-wellbeing-indicator.pdf (accessed 31/01/2024).

Law, J., Rush, R., King, T., Westrupp, E. and Reilly, S. (2018) 'Early Home Activities and Oral Language Skills in Middle Childhood: A Quantile Analysis'. *Child Development*, 89: 1, 295–309. https://doi.org/10.1111/cdev.12727

Law, J., Rush, R., Schoon, I. and Parsons, S. (2009) 'Modeling Developmental Language Difficulties From School Entry into Adulthood: Literacy, Mental Health, and Employment Outcomes'. *Journal of Speech, Language, and Hearing Research*, 52: 6, 1401–1416. https://doi.org/10.1044/1092-4388(2009/08-0142)

Masset, E. and A. Butt (2025) Efficacy Trial of Talking Time, an Oral Language Intervention for Early Years: Statistical Analysis Plan, EEF, https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/talking_time_-_statistical_analysis_plan_-_v.1.0.pdf?v=1782315127

Massonnié, J., Llaurado, A., Sumner, E. and Dockrell, J.E. (2022) 'Oral Language at School Entry: Dimensionality of Speaking and Listening Skills'. *Oxford Review of Education,* 48: 6, 743–766. https://doi.org/10.1080/03054985.2021.2013189

Mathers, S., Malmberg, L.E., Lindorff, A. and Gardiner, J. (2022) *'Early Educators' Knowledge of Early Language Pedagogy: How Can It Be Measured and Does It Matter for Child Outcomes'*. London: Nuffield Foundation. Available at: Measuring pre-school teachers' knowledge of oral language pedagogy - Nuffield Foundation (accessed 01/05/2026).

Michie, S., van Stralen, M.M. and West, R. (2011) 'The Behaviour Change Wheel: A New Method for Characterising and Designing Behaviour Change Interventions'. *Implementation Science*, 6: 42. https://doi.org/10.1186/1748-5908-6-42

Morra Pellegrino, M.L. and Scopesi, A. (1990) 'Structure and Function of Baby Talk in a Day-Care Centre'. *Journal of Child Language*, 17: 1, 101–114. https://doi.org/10.1017/S030500090001312X

Muñoz, J., Hufstedler, H., Gustafson, P., Bärnighausen, T., De Jong, V.M.T. and Debray, T.P.A. (2023) 'Dealing with Missing Data Using the Heckman Selection Model: Methods Primer for Epidemiologists'. *International Journal of Epidemiology*, 52: 1, 5–13. https://doi.org/10.1093/ije/dyac237

Nelson, K.E., Welsh, J.A., Trup, E.M.V. and Greenberg, M.T. (2011) 'Language Delays of Impoverished Preschool Children in Relation to Early Academic and Emotion Recognition Skills'. *First Language*, 31: 2, 164–194. https://doi.org/10.1177/0142723710391887

Ofsted. (2022) *'Education Recovery in Early Years Providers: Summer 2022'*. GOV.UK. Available at: www.gov.uk/government/publications/education-recovery-in-early-years-providers-summer-2022/education-recovery-in-early-years-providers-summer-2022 (accessed 31/01/2024).

Pace, A., Luo, R., Hirsh-Pasek, K. and Golinkoff, R.M. (2017) 'Identifying Pathways Between Socioeconomic Status and Language Development'. *Annual Review of Linguistics,* 3, 285–308. https://doi.org/10.1146/annurev-linguistics-011516-034226

Renfrew, C. (2010) '*The Bus Story'*. Chesterfield: Winslow Resources. Available at: https://www.winslowresources.com/bus-story-test-renfrew-language-scales.html (accessed 31/01/2024).

Renfrew, C. (2019) '*Renfrew Action Picture Test (5th Edition)'*. Milton Keynes: Speechmark Publishing Ltd.

Renfrew, C. (2023) '*Renfrew Expressive Vocabulary Test (5th Edition)'*. Milton Keynes: Speechmark Publishing Ltd.

Romeo, R.R., Leonard, J.A., Robinson, S.T., West, M.R., Mackey, A.P., Rowe, M.L. and Gabrieli, J.D.E. (2018) 'Beyond the 30-Million-Word Gap: Children's Conversational Exposure Is Associated With Language-Related Brain Function'. *Psychological Science*, 29: 5, 700–710. https://doi.org/10.1177/0956797617742725

Roulstone, S., Law, J., Rush, R., Clegg, J. and Peters, T. (2010) *'Investigating the Role of Language in Children's Early Educational Outcomes'*. London: Department for Education.

Rowe, M.L. (2022) 'Environmental Influences on Early Language and Literacy Development: Social Policy and Educational Implications'. *Advances in Child Development and Behavior,* 63: 103–127. https://doi.org/10.1016/bs.acdb.2022.04.001

Rowe, M.L. and Snow, C.E. (2020) 'Analyzing Input Quality Along Three Dimensions: Interactive, Linguistic, and Conceptual'. *Journal of Child Language*, 47: 1, 5–21. https://doi.org/10.1017/S0305000919000655

Rubin, D.B. (1976) 'Inference and Missing Data'. *Biometrika*, 63: 3, 581–592. https://doi.org/10.1093/biomet/63.3.581

Sims, S., Fletcher-Wood, H., O'Mara-Eves, A., Cottingham, S., Stansfield, C., Van Herwegen, J. and Anders, J. (2021). *'What Are the Characteristics of Effective Teacher Professional Development? A Systematic Review and Meta-Analysis'*. London: Education Endowment Foundation. Available at: https://educationendowmentfoundation.org.uk/education-evidence/evidence-reviews/teacher-professional-development-characteristics (accessed 01/05/2026).

Strand, S., Malmberg, L., and J. Hall (2015) *'English as an Additional Language (EAL) and Educational Achievement in England: An Analysis of the National Pupil Database'*. Cambridge and Oxford: The Bell Foundation and University of Oxford. Available at: www.bell-foundation.org.uk/app/uploads/2025/11/EAL-attainment-2025-Oxford-report.pdf (accessed 11/02/2026).

Van Buuren, S. (2018) *Flexible Imputation of Missing Data (2nd Edition)'*. New York, NY: CRC Press.

van der Lely, H.K.J., Payne, E. and McClelland, A. (2011) 'An Investigation to Validate the Grammar and Phonology Screening (GAPS) Test to Identify Children with Specific Language Impairment'. *PLoS ONE*, 6: 7, e22432. https://doi.org/10.1371/journal.pone.0022432

Wasik, B. (2008) 'When Fewer is More: Small Groups in Early Childhood Classrooms'. *Early Childhood Education Journal*, 35: 6, 515–521. https://doi.org/10.1007/s10643-008-0245-4

Wright, H., Carr, D., Wiese, J., Stokes, L., Runge, J., Dorsett, R., Heal, J. and Anders, J. (2020) *'URLEY (Using Research Tools to Improve Language in the Early Years): Evaluation Report'*. London: Education Endowment Foundation. Available at: https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/using-research-tools-to-improve-language-in-the-early-years (accessed 07/02/2024).

Appendix A: The EEF cost rating

Appendix A Table 1: Cost rating

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

Appendix B: Security classification of trial findings

OUTCOME: Composite standardised index of Expressive vocabulary (REV) and of Information and Grammar (RAPT)

Secondary outcome analysis and/or subgroup analyses are NOT included in the security ratings unless otherwise stated.

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

Threats to validity	Risk rating	Comments
Threat 1: Confounding	Low	Randomisation and restricted allocation achieved good baseline balance on key characteristics; analysis adjusts for baseline and stratification.
Threat 2: Concurrent Interventions	Low	Both IPE interviews and surveys indicate language-focused activity across settings beyond Talking Time. In terms of staff survey, evidence suggests that about 50% of settings had some form of concurrent intervention at baseline (85% response rate). However, we consider the threat to internal validity to be low to moderate. As shown in Table 13 of the report, arm-level evidence on exposure to concurrent provision was balanced between intervention and control groups.
Threat 3: Experimental effects	Low	Assessors blinded, no indications of differential assessment practice. Important to flag child-level attrition, slightly higher in intervention, but overall balanced.
Threat 4: Implementation fidelity	Low	Fidelity generally strong; some variability in dosage and sequencing, but no clear differential by arm.
Threat 5: Missing Data	Low/Moderate	Primary-outcome analytical sample requires completion of multiple assessments; overall missing data 15.6% (moderate). Missingness is associated with EYPP/EAL, but comparisons and logistic models show no differential missingness by arm and robustness checks are stable. Table 22 shows that the number of missing observations is very similar across the

		Renfrew assessments: 174 for the combined index, 179 for REV, and 161 for RAPT.
Threat 6: Measurement of Outcomes	Low/Moderate	Equal-weight arithmetic mean used after PCA suggested near-equal weights. The composite index has been constructed using standardised scores, addressing concerns whether the arithmetic mean could skew towards REV (0–100 scale). Report provides rationale and robustness checks.
Threat 7: Selective reporting	Low	Outcomes reported per SAP; deviations explained. Analytical approach was identified before outcomes were observed,

- **Initial padlock score:** 4 Padlocks – Due to attrition.
- **Reason for adjustment for threats to validity**: no adjustments. The trial is a well-conducted cluster RCT with adequate power and balanced attrition. On this basis we provide a 4-padlock rating.
- **Final padlock score:** initial score adjusted for threats to validity = 4 Padlocks.
- **Reason(s) for final padlock rating:** See above.

Further Appendices

Appendix C to Appendix J can be found published separately on the project page, labelled as ‘Further Appendices’.

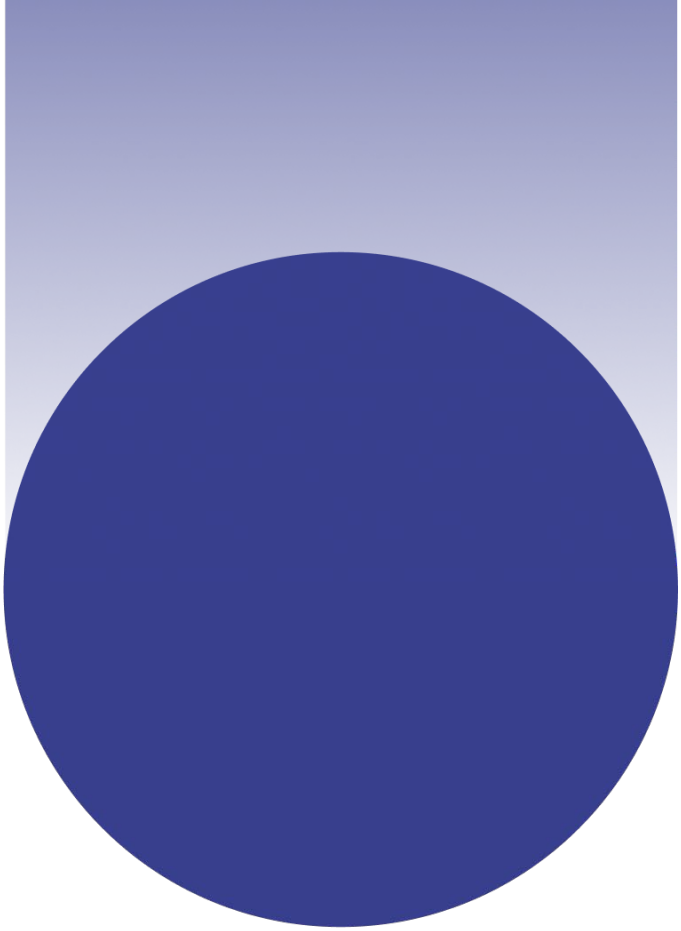
You may re-use this document/publication (not including logos) free of charge in any format or medium, under the terms of the Open Government Licence v3.0.

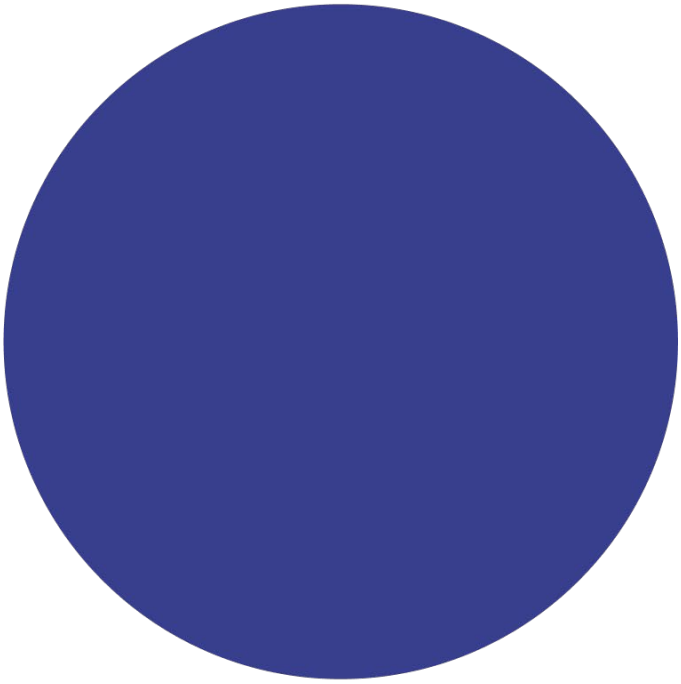
To view this licence, visit https://nationalarchives.gov.uk/doc/open-government-licence/version/3 or email: psi@nationalarchives.gsi.gov.uk

Where we have identified any third-party copyright information you will need to obtain permission from the copyright holders concerned. The views expressed in this report are the authors' and do not necessarily reflect those of the Department for Education.

This document is available for download at https://educationendowmentfoundation.org.uk







Education Endowment Foundation

The Education Endowment Foundation
5th Floor, Millbank Tower,
21–24 Millbank,
London,
SW1P 4QP

https://educationendowmentfoundation.org.uk

@EducEndowFoundn

Facebook.com/EducEndowFoundn