

Independent evaluation of the Early Years Conversation Project (EYCP): A two-armed cluster randomised waitlist-controlled trial

Evaluation report

Xiaofei Qi, Akansha Singh, Rachel Oughton, Laura Chambers, Rowan Van-Muysen, Dandan Chen, Yuting Guo, Nadia Siddiqui, and Victoria Menzies

October 2026



The Education Endowment 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 2–19-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 re-endowed with an additional £137m from the 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
SW1P 4QP

0207 802 1653

info@eefoundation.org.uk

www.educationendowmentfoundation.org.uk

Table of Contents

Executive summary **4**

Introduction **6**

Methods **17**

Impact evaluation results **43**

Implementation and process evaluation results **59**

Conclusion **113**

References **119**

Appendix A: EEF cost rating **123**

Appendix B: Security classification of trial findings **124**

Appendix C: Effect size estimation **126**

Further appendices **127**

About the evaluator

The project was independently evaluated by a team from Durham University: Xiaofei Qi, Vic Menzies, Rachel Oughton, Nadia Siddiqui, Akansha Singh, Laura Chambers, Dandan Chen, and Rowan Van-Muysen. The project was jointly led by Xiaofei Qi and Vic Menzies.

Contact details

Xiaofei Qi

School of Education, Durham University

Confluence Building

Stockton Road

Durham, DH1 3LE

Email: **xiaofei.qi@durham.ac.uk**

Acknowledgements

We would like to acknowledge and thank the nursery settings and practitioners who took part in this programme along with the children and families involved. We would like to thank the delivery team at East London Research School for their cooperation in all aspects of this evaluation; it has been a pleasure to work with you all. We would also like to acknowledge our team of language assessors at Durham who travelled all across London to each of the settings to work with the participating children in assessing their language outcomes at different timepoints. We would like to acknowledge the work of Sandra Mathers and her team at Oxford University for the provision and scoring of the Observing Language Pedagogy (OLP) measure.

Finally, we would like to thank Ashley High and Alex Jones in our Research Administration team for their help and support throughout the project.

Executive summary

The project

The Early Years Conversation Project (EYCP) is an evidence-informed professional development (PD) programme developed and delivered by the East London Research School at Sheringham Nursery School and Children's Centre. It aims to improve the communication and language skills of children aged two to three by supporting early years practitioners to develop their knowledge and use the ShREC approach (Share attention, Respond, Expand, and Conversation) in everyday interactions. The EYCP programme runs for eight months, consisting of one and a half days of in-person training and six hours of live online modules, specialist mentoring by experienced Early Years mentors through small group hub meetings and setting support visits, and access to a bank of online resources. The trained EYCP Champion works to implement the EYCP strategies in their setting, supported by their nursery lead or manager.

The efficacy evaluation used a cluster randomised trial design involving 102 nursery settings across the Greater London area with 902 children. Settings were randomly assigned to receive the EYCP programme or to receive EYCP the following year (waitlist-control group). The trial was delivered over two cohorts with the first cohort delivered in 2023/2024 and the second cohort delivered in the 2024/2025 school year (October–July). An implementation and process evaluation used a mixed-methods design combining interviews, focus groups, surveys, observations of training and hub meetings, audits of settings' learning environment, and administrative records to assess how EYCP was delivered in practice.

As part of the Department for Education (DfE)'s Early Years Recovery Programme (DfE, 2021), the Education Endowment Foundation (EEF) has worked with Stronger Practice Hubs across England to fund early years settings' access to evidence-informed programmes and study the programme's influence on practice and children's outcomes. This initiative aimed to support education recovery following the pandemic while also developing understanding of effective professional development in the early years. The EEF has worked with 'A Brighter Start' Stronger Practice Hub to fund settings' access to EYCP and to evaluate the programme.

Table 1: Key conclusions

Key conclusions
1. Children in EYCP intervention nurseries made no additional months' progress in receptive and expressive vocabulary, on average, compared to children in the waitlist-control nurseries. This result has a low to moderate security rating.
2. Among children living in the areas of highest deprivation, those in EYCP intervention nurseries made two fewer months progress in receptive vocabulary, on average, compared to those in waitlist-control nurseries. These results may have lower security than the overall findings because of the smaller number of children.
3. Children in nurseries that implemented EYCP with high fidelity and who attended the setting regularly appeared to make more progress in receptive vocabulary compared to children in control nurseries. This result is less secure than the main finding due to the smaller sample size.
4. The EYCP programme improved practitioners' language-related pedagogical knowledge. This finding has low security because of the small number of practitioners.
5. Practitioners reported increased confidence in supporting children's language and communication, along with changes in classroom practice through implementation of the EYCP strategies using the ShREC approach and interactive book reading. The programme was perceived to benefit all children, regardless of their language background.

EEF security rating

These findings have a low to moderate security rating. This trial was an efficacy trial, which tested whether the intervention worked under developer-led conditions in a number of schools. The trial was a well-designed split cohort, cluster randomised waitlist-controlled trial. The trial was well powered and balanced across intervention and control arms in terms of children's characteristics. However, there was a slight imbalance in baseline test scores across intervention and control arms. In addition, 21.6% of the children who started the trial were not included in the final analysis because they did not complete endline testing, with more data missing for children in the intervention group than the control group. These factors make it harder to accurately estimate the size of the impact on the children in the trial.

Additional findings

Children in EYCP nurseries made, on average, no additional progress in their expressive and receptive vocabulary and language and communication skills compared to those in the waitlist-control group equivalent. This is our best estimate of impact, which has a low to moderate security rating. As with any study, there is always some uncertainty around the result: the confidence interval was wide, encompassing effects ranging from around three months' less progress to two months' additional progress on the primary outcome, receptive vocabulary. Combined with the EEF security rating of two, this suggests that the findings should be interpreted as inconclusive rather than evidence that the programme is ineffective.

Implementation and process evaluation findings showed strong nursery staff engagement with and satisfaction in EYCP. The EYCP Champion role effectively supported delivery, although manager and EY lead involvement was more variable. Key barriers to engagement included staff release, workload, travel to training and mentoring, and challenges of working with two-year-olds, while facilitators included mentor support, accessible resources, leadership buy-in, and video-based reflection.

While the programme improved practitioner knowledge and practice, it did not produce detectable improvements in children's language outcomes within the study timeframe. This pattern is consistent with wider evidence showing that professional development programmes can improve practitioners' behaviour, but that impacts on language outcomes for children aged two to three are often small, variable, or take longer to emerge (Dockrell *et al.*, 2023; Markussen-Brown *et al.*, 2017). This trial highlights the challenges of achieving measurable language gains among children aged two to three years within a relatively short follow-up period. However, the current trial cannot determine whether longer-term follow-up would have detected stronger effects. Encouragingly, stronger effects among children exposed to higher-fidelity implementation and greater engagement suggest the programme has promise. Further implementation support, stronger leadership involvement, and longer-term follow-up may help translate practice improvements into child-level gains.

Cost

The average cost of EYCP for a nursery setting was around £2,972 for three years of delivery, or £47 per child per year when averaged over three years. This was using the Cohort 2 delivery model with eight mentors supporting 49 settings. EYCP required 36 hours of practitioner and 11 hours of senior leader/management time to engage with training and support over the course of the project. As EYCP is delivered at a setting level, the cost per child varies significantly depending on the number of children in the setting, implying that smaller settings would have a higher rate per child compared to larger settings. Although EYCP strategies are incorporated into everyday nursery practice, ongoing practitioner preparation and planning time was estimated at 1.5 hours per week which would be assumed to continue into subsequent years.

Impact

Table 2: Summary of impact on primary outcome(s)

Outcome/ Group	Effect size (95% confidence interval)	Estimated months' progress	EEF security rating	No. of children	*P*-value	EEF cost rating
Receptive vocabulary - All children	-0.03 (-0.24, 0.18)	0		707	0.77	£ £ £ £ £
Receptive vocabulary - Children living in areas of highest deprivation	-0.11 (-0.43, 0.22)	-2	N/A	264	0.56	£ £ £ £ £

Introduction

Background

Early language ability is one of the strongest predictors of later development through school and in life (Burchinal *et al.*, 2016; Hoff, 2013; Pace *et al.*, 2017). Language is essential for thinking, social, and emotional well-being and learning in all areas of development. The significance of early language development and communication skills has been widely recognised in the Early Years Foundation Stage (EYFS) (Department for Education, 2023a). Language and communication in the early years are fundamental skills for future learning and can predict language outcomes and reading skills throughout primary school; the persistence of low language skills is linked with lower academic achievement (Snow, Burns, and Griffin, 1998; Eadie *et al.*, 2021; Sénéchal, Ouellette, and Rodney, 2006).

Influences such as the home learning environment and the preschool learning environment have been shown to play essential roles in children's development at early ages (Lehrl, Evangelou, and Sammons, 2020; Melhuish *et al.*, 2008; Sylva *et al.*, 2004). Existing evidence shows that providing children with the experience of a varied and rich language and literacy exposure at home and in preschool provision, such as sharing books, conversations, and using child-directed speech, promotes children's language and literacy development (Mol and Neuman, 2014; Hoff, 2013; Law *et al.*, 2017a).

It is well documented that the language skills of preschool children differ substantially and that these differences are highly predictive of their later academic success and achievements. The prevalence of language delays and difficulties in early years is of great concern (Public Health England, 2020a), and it is argued that cost-effective, evidence-based training and interventions that promote the most effective types of language-boosting interactions between children and those caring for them (parents/carers and early years practitioners) are needed to ensure that all children have the best possible chance of reaching their full potential (Law *et al.*, 2017b; Wooles, Swann, and Hoskison, 2018).

Many children from disadvantaged backgrounds have poorer communication, language, and vocabulary (Law *et al.*, 2017b; Wooles, Swann, and Hoskison, 2018). There is evidence of language gaps for disadvantaged children at the beginning of school; children from low-income backgrounds consistently perform below their more advantaged peers on standardised measures of language ability, setting long-term trajectories that translate into gaps in academic achievement (Pace *et al.*, 2017; Schwab and Lew-Williams, 2016). According to statistics reports on government-funded early years provision for children under 5 years of age in England, most disadvantaged children are in private, voluntary, or independent (PVI) nursery settings. For example, 625,434 children under age 4 are registered for 15-hour entitlement with PVI providers in 2023, and only 298,903 children are registered with state-funded schools (see Education provision: children under 5 years of age; GOV.UK, 2023a). However, staff in PVIs have lower levels of qualifications and less access to professional development (Haux *et al.*, 2022), including professional development focused on communication and language, when compared to staff in schools. The funded two-year-old entitlement specifically targets children from disadvantaged backgrounds who are at increased risk of poorer developmental outcomes, including delays in communication and language. Consequently, interventions delivered within settings serving these children have the potential to reduce inequalities before school entry and improve longer-term educational trajectories.

In the EEF Early Years Toolkit: Communication and language approach (EEF, n.d.), a small number of studies conducted in nurseries serving higher proportions of socio-economically disadvantaged children have found above-average effects on those children's learning outcomes, suggesting that targeted language and communication support is likely to be a beneficial approach for this group to narrow inequalities and learning gaps before school age (Law *et al.*, 2017b). Communication and language approaches that emphasise the importance of spoken language and verbal interaction for young children have shown promising results in the EEF Early Years Toolkit (EEF, n.d.). EYCP (formerly piloted as Manor Park Talks), developed and delivered by East London Research School (ELRS), is a PD programme designed to support early years/nursery staff with developing and understanding practical strategies to enhance children's communication and language skills. The programme draws on a broader evidence base demonstrating that communication and language approaches benefit young children's learning (EEF, n.d.). EYCP aims to promote conversational responsiveness by training early years professionals to implement and embed the 'ShREC approach' (Share attention, Respond, Expand, and

Conversation) in everyday practice (Ang *et al.*, 2025). Recent work by Ang and colleagues (Ang *et al.*, 2025) argues that ShREC operationalises evidence on conversational responsiveness into a practical and scalable pedagogical model for early years educators, particularly through multi-turn conversations, responsive interaction, scaffolding, and interactive book reading. The approach aims to embed these interactional practices into everyday nursery routines through intentional pedagogy and reflective practice.

As part of the Department for Education's (DfE) Early Years Recovery Programme, A Brighter Start Stronger Practice Hub and the Education Endowment Foundation (EEF) are working together to fund early years settings with access to the EYCP. An Evaluation Team (ET) from the University of Durham has been appointed to evaluate the impact of EYCP on young children's communication and language development.

An EEF-funded pilot evaluation of the programme was conducted by the National Centre for Social Research (NatCen), across 12 early years settings and 6 schools in Newham (Husain, Morris, Basi, and Bristow, 2020). Findings from the pilot demonstrated evidence of promise for improving children's communication and language outcomes, as hypothesised in the programme's theory of change (ToC). The pilot results also indicated that the strategies were becoming embedded in daily practice. Given this context and existing evidence, we developed an efficacy trial that integrated a randomised controlled trial (RCT) to evaluate the impact of the EYCP on 2–3-year-old nursery children's language and communication skills with an implementation and process evaluation (IPE) to explore the causal pathways of the ToC and identify elements of effective delivery. (See Table 3 for evaluation design.) The efficacy trial allowed us to rigorously investigate the potential impact of the programme under best possible conditions. It also adds to the evidence base on the benefits of a language programme that supports staff working with 2–3-year-olds, an area where there is currently an evidence gap.

Intervention

The following description of the Early Years Conversation Project (EYCP) has been structured using the framework for the Template for the Intervention Description and Replication (TIDieR) checklist as shown in the EYCP Evaluation Protocol report **[here]**.

Why: Rationale, theory, and/or goals of essential elements of the intervention

Background

The rationale for the EYCP is grounded in evidence showing that a key determinant of language development is the amount and quality of language to which a young child is exposed. Many children from disadvantaged backgrounds have poorer communication, language, and vocabulary, and most of these children attend private, voluntary, or independent (PVI) nursery settings. Staff in PVIs typically have lower qualification levels and less access to professional development, including training in communication and language, compared with staff in schools.

Overview

EYCP focuses on conversational responsiveness, using adult-child interactions to increase both the amount children say and the complexity of their conversations. Its aim is to improve the skills and expertise of the early years workforce to better implement techniques that promote conversational responsiveness and support children's language development. PD for practitioners in early years settings is based around the EEF's ShREC approach. The ShREC approach includes four evidence-informed strategies: Share attention, Respond, Expand, and Conversation. These strategies are intended to be used throughout the child's day in nursery, with an emphasis on multi-turn conversations, decontextualised talk, and vocabulary. Interactive book reading is a key vehicle for this approach.

Who: Recipients of the intervention

The recipients of the intervention include PVI nurseries, state-funded school-based nurseries, and maintained nurseries offering Free Early Education (FEE) and childcare for two-year-olds in selected London local authorities (see Where section for more details). Two practitioners from each setting participate: the setting manager or Early Years Foundation Stage (EYFS) lead, and the EYCP Champion, who is a room lead or nursery teacher working with two- to three-year-olds. Although

the programme aims to improve language and communication provision for all children in participating nurseries, the direct recipients are the two- to three-year-old children in the EYCP Champion's classroom.

What: Physical or informational materials used in the intervention

Professional development is delivered through a mix of face-to-face and virtual training, consisting of one and a half days of in-person training and six hours of live online modules spreading across the whole intervention year between November and June. Online modules will be recorded, with recordings hosted on an online platform such as Padlet. Specialist mentoring is provided through a combination of face-to-face and virtual hub meetings and support visits from a mentor. Participants also have access to a dedicated online knowledge base, which includes video exemplification, background materials, and resources to support families at home. A physical handbook with the programme timetable and photocopiable resources is also provided.

What: Procedures, activities, and/or processes used in the intervention

PD training is delivered through both face-to-face and virtual sessions. The EYCP Champion receives all training, while the setting manager or EYFS leader attends only specific sessions. EYCP mentors, who were experienced early years specialists recruited by East London Research School as part of the delivery team, provide specialist support through nine hub meetings across the year and a minimum of five support visits. Mentors will mentor a specified number of settings during the intervention period. If possible, the same group of mentors (n = 8) will stay in the EYCP programme over three years: in the 2023–2024 school year (Cohort 1), eight mentors will work in pairs, each pair to support 7 settings; in the 2024–2025 school year (Cohort 2), each mentor will independently support 7 settings; and in the 2025–2026 school year, each mentor is expected to support 7 settings (amended or protocol v2.0). The difference in mentor models between Cohort 1 and Cohort 2 is due to the split cohort trial design following recruitment challenges. The pairing of mentors allowed the retention of mentors who were already recruited and provided additional support to mentors for their first year of delivery.

At the beginning and end of the programme, mentors complete an ITERS-3 (Infant/Toddler Environment Rating Scale) (Harms *et al*., 2017) observation as part of the intervention in the setting to provide formative feedback on the setting's progress towards providing a language-rich environment. Hub meetings will be used to build knowledge, revisit online learning, embed understanding, and support implementation. Mentors use video to help participants recognise effective strategy use and its impact on children. These meetings will provide needs-based support for the EYCP Champion to implement the strategies, overcome barriers, and cascade strategies to the wider team. EYCP Champions are expected to implement the ShREC strategies—Share attention, Respond, Expand, and Conversation— with young children in everyday practice.

Who: Intervention providers/implementers

The PD training is delivered by Fliss James and Melissa Prendergast (referred to here as the ELRS team), supported by a team of eight mentors who provide in-setting support and run hub meetings. Mentors are trained by the ELRS team prior to the start of the intervention. Training includes essential evidence, strategies, and features of effective PD. Mentors are trained in the use of ITERS-3 and are moderated. Mentors attend all PD training alongside practitioners and meet regularly with the ELRS team.

The EYCP Champion implements the strategies daily with all children in their classroom and may cascade training to their wider team through an informal buddy system where the trained EYCP Champion partners with an untrained other member of staff to train and support their buddy in implementing the strategies. Use of the informal buddy system varied across settings and was not a mandatory component of the programme. The setting manager engages with PD on programme implementation and supports the Champion in embedding the strategies.

How: Mode of delivery

EYCP includes a mix of online and face-to-face PD training and mentoring sessions, with PD sessions, hub meetings, and support visits running in an alternating cycle.

Where: Location of the intervention

EYCP is available to nursery settings across Greater London[1] from the following local authorities, including Barking and Dagenham, Barnet, Bexley, Brent, Bromley, Camden, City of London, Croydon, Ealing, Enfield, Greenwich, Hackney, Hammersmith and Fulham, Haringey, Harrow, Havering, Hillingdon, Hounslow, Islington, Kensington and Chelsea, Kingston Upon Thames, Lambeth, Lewisham, Merton, Newham, Redbridge, Richmond Upon Thames, Southwark, Sutton, Tower Hamlets, Waltham Forest, Wandsworth, and Westminster (amended for protocol v2.0).

When and how much: Duration and dosage of the intervention

EYCP lasts eight months and is delivered across three school years: 2023–2024 (Cohort 1), 2024–2025 (Cohort 2 and Cohort 1 waitlist delivery), and 2025–2026 (Cohort 2 waitlist delivery).

Setting managers attend one and a half days of face-to-face training and one online session. EYCP Champions attend all training: one and a half days face-to-face, six online sessions, and nine hub meetings. They receive one mentor support visit per half term, with additional visits if needed (one hour; focus on content and behaviour change). Champions are expected to use EYCP strategies every day with every child in their classroom.

Tailoring: Adaptation of the intervention

Although the PD training and mentoring are manualised, bespoke support is provided based on participant need. All aspects of the intervention are evaluated and monitored to identify where bespoke support might be needed. PD sessions delivered by the ELRS team are evaluated, and any misconceptions or gaps in knowledge are responded to by meeting with mentors and supporting them to address these in hub meetings and support visits. Mentors reflect on their work, and their self-evaluations are shared with the ELRS team. The ELRS team and mentors meet regularly to coordinate tailored support. The ELRS team have oversight of this tailoring allowing for fidelity/adherence control and checks.

How well (planned): strategies to maximise effective implementation

The ELRS team draws on experience from three previous iterations of the programme, allowing for an extensive exploration and preparation stage. The PD programme is manualised and aligned with the EEF's guidance on professional development, incorporating mechanisms to build knowledge, motivate staff, develop teaching techniques, and embed practice. Mentors access training materials via Padlet, including mentor-specific resources and reflection documents monitored by the ELRS team. Mentors work closely with participants to ensure successful implementation of EYCP strategies. The ELRS team also supports participants to scale and sustain the strategies through a PD session on implementation and ongoing mentor and peer support.

A more specific logic model was developed for this efficacy trial (see Figure 1). The causal mechanism of the logic model is that providing PD for practitioners in early years settings to use and understand effective language and communication strategies based around the 'ShREC approach' (Ang *et al.*, 2025) can enhance practitioners' language-related pedagogical knowledge and improve children's vocabulary, language, and communication skills. The targeted population are 2–3-year-olds in PVI, maintained nurseries, and SN settings; setting managers/EYFS lead and EYCP Champions. The detailed EYCP intervention inputs/activities, outputs and expected outcomes for the targeted population can be seen in Figure 1.

[1]Nursery settings were recruited from the Greater London area because the East London Research School, which developed the EYCP programme and oversees delivery, is based in this region. This geographical focus enabled more efficient coordination of programme delivery, school engagement, and hub meetings during the trial.

Figure 1: Logic model of EYCP efficacy trial

Causal mechanism: Increasing early years practitioners' use and understanding of effective language and communication strategies based around the 'ShREC approach' will enhance practitioner's language-related pedagogical knowledge and improve children's vocabulary, language, and communication skills.

Inputs/Activities	Outputs	Outcomes
Setting managers/EYFS leads: • PD training through a mix of face to face (1.5 days) and virtual session (1 session) (Core activity). • Access to EYCP programme handbooks dedicated online knowledge base (Optional activity). • Engage with online networking activity and programme support (Optional activity). **EYCP Champions:** • PD training through a mix of face to face (1.5 days) and virtual sessions (6 pre-recorded sessions) (Core activity). • Specialist mentoring through a mix of face to face and virtual 'hub' meetings (9 hub meetings, minimum 5 in-setting visits and 2 IERS-3 audits from mentors) (Core activity). • Access to EYCP handbooks and dedicated online knowledge base (Core activity). • Strategies cascaded from mentor to EYCP Champion via modelling, videoing, and reflection (Core activity). • Engage with online networking activity and programme support (Optional activity). **Children:** Experience a richer language environment, with higher quality early years teaching involving regular multi-turn conversations through the use of the ShREC approach, including interactive book reading.	**Setting managers/EYFS leads:** • Engage with PD training and robust evidence about communication and language. • Support EYCP Champions' behaviour change and EYCP implementation. **EYCP Champions:** • Engage with robust evidence about communication and language. • Engage with PD, mentoring, and in-setting support visits from mentors. • Engage with online networking opportunities and programme support. • Engage with children using the 'ShREC Approach'. • Cascades strategies to their wider team. **Children:** Experience practitioners sharing attention with them; engage with practitioners in sustained contingently responsive conversation; hear increasingly complex language.	**Primary outcome: (impact)** • Improvement in children's receptive and expressive vocabulary attainment. **Secondary outcomes: (impact)** • Improvement in children's language and communication skills. • Improvement in practitioners' language-related pedagogical knowledge. **Exploratory outcomes: (process and impact)** • Improvement in practitioner confidence in language teaching. • Observed changes of language provision learning environment.

Potential mediators/moderators: Children: FEE and EAL status at age 2, Baseline performance on vocabulary attainment; Practitioner: teaching qualification and teaching experience.

Evaluation objectives

The latest version of the EYCP evaluation protocol and statistical analysis plan (SAP) for the current efficacy trial can be found on the EEF project website [**here**].

Research questions

In the impact evaluation, our primary research question is:

- What is the impact of the EYCP on the receptive vocabulary [measured by the BPVS-III] of 2–3-year-olds in intervention nurseries receiving the EYCP in comparison to those children in the waitlist-control nurseries receiving business as usual? *(Primary outcome)*

Our secondary impact evaluation research questions are:

- What is the impact of the EYCP on the expressive vocabulary [measured by the CDI-III] of 2–3-year-olds in intervention nurseries receiving the EYCP in comparison to those children in the waitlist-control nurseries receiving business as usual? *(Secondary outcome)*
- What is the impact of the EYCP on the language and communication skills [measured by the WellComm Early Years toolkit] of 2–3-year-olds in intervention nurseries receiving the EYCP in comparison to those children in the waitlist-control nurseries receiving business as usual? *(Secondary outcome)*
- What is the impact of the EYCP on practitioners' language-related pedagogical knowledge [measured by the OLP tool] in intervention nurseries receiving the EYCP in comparison to those practitioners in the waitlist-control nurseries receiving business as usual? *(Secondary outcome)*
- What is the impact of the EYCP on the receptive and expressive vocabulary of children who are from disadvantaged backgrounds (represented by IDACI 1–3[2]) in intervention nurseries receiving the EYCP in comparison to the subgroup of children in the waitlist-control nurseries receiving business as usual? *(Primary and secondary outcome: disadvantaged background subgroup)*
- What is the impact of the EYCP on the receptive and expressive vocabulary of children with English as an additional language (EAL) in intervention nursery settings receiving the EYCP in comparison to that subgroup of children in the waitlist-control nurseries receiving business as usual? *(Primary and Secondary outcome: EAL subgroup)*
- What is the impact of the EYCP on the receptive and expressive vocabulary of children with lowest/best baseline performance [measured by the BPVS-III and ELIM] in intervention nurseries receiving the EYCP in comparison to that subgroup of children in the waitlist-control nurseries receiving business as usual? *(Primary and secondary outcome: best/lowest performance subgroup)*

Our main IPE research questions (RQ) and the subgroup questions are listed below.

Fidelity/adherence

RQ1: Is the EYCP PD training and mentoring support delivered as intended to PVI, maintained, and SN settings?

1.1. Are nominated staff (manager/EYFS lead, EYCP Champion) attending EYCP training and the hub meetings (EYCP Champion) as specified in the programme plan?

[2] At the time of the study in 2023, only 3- and 4-year-old children were eligible for the Early Years Child Premium (EYPP). The use of Free Early Education (FEE) entitlement eligibility and the Income Deprivation Affecting Children Index (IDACI) was therefore specifically intended to identify disadvantage within the two-year-old age group. This approach aligned with the government's early years funding consultation proposals at the time.

1.2. How appropriate are the PD training sessions (for example, content, coverage, dosage, and duration) and the level of support for nominated staff? Are PD training sessions and mentoring support (including hub meetings and setting visits) delivered as manualised?

1.3. What are the barriers and required facilitators for nominated staff to access and engage with the EYCP PD training and mentoring sessions online and/or face to face (including hub meetings and setting visits)?

1.4. Is the EYCP PD training and support delivered to settings consistently by EYCP mentors across different hubs?

1.5. To what extent does the EYCP programme and its delivery to settings in Cohort 1 differ from the delivery in Cohort 2?

RQ2: Is the EYCP programme implemented with fidelity in the intervention group in PVI, maintained, and SN settings?

2.1. What are the perceived roles of nominated staff (manager/EYFS lead, EYCP Champion) in the programme?

2.2. Do nominated staff adhere to their roles as requested in the programme?

2.3. To what extent are the EYCP strategies cascaded to the wider team within the setting?

2.4. What are the barriers and facilitators for implementing the programme in PVI, maintained, and SN settings?

2.5. What other factors explain variation in the fidelity of implementation in the intervention group?

2.6. To what extent does the EYCP implementation in Cohort 1 settings differ from the implementation in Cohort 2 settings?

Usual practice

RQ3: What is the usual practice in language and communication provision like in all nursery settings?

3.1. What are the key observable features of a setting delivering the EYCP (in terms of interactions with children, etc.) in the intervention group?

3.2. To what extent do the EYCP features differ from the usual language and communication practice in the control group?

RQ4: What is the usual PD practice like in all nursery settings?

4.1. What language-related PD opportunities do staff have in the control group (for example, DfE's Early Years Education Recovery Programme's Online Early Years Child Development Training strand)?

4.2. What are the perceived impacts of these language-related PD programmes on children's language and communication skills in the control group?

4.3. To what extent does the EYCP PD programme differ from the usual PD practice in all settings?

4.4. Are nursery staff participating in any other language-related PD programmes (for example, DfE's Early Years Education Recovery Programme's Online Early Years Child Development Training strand) in the intervention group?

Causal mechanism and contextual assumptions

RQ5: What are the perceived impacts of the EYCP from different stakeholders (including developers, mentors, setting managers, EYCP Champions, and practitioners)?

5.1. How far does the EYCP PD training and mentoring support impact the knowledge, skills, and confidence of EYCP Champions regarding children's language and communication needs?

5.2. What are the observed impacts of the EYCP upon the setting's learning environment as evidenced by the ITERS-3 audit outcome?

5.3. What are the observed impacts of the EYCP upon practitioners' language teaching practice and skills?

5.4. What are the observed impacts of the EYCP on children's language learning as well as other learning areas in the nursery?

5.5. What factors appear to affect the intervention impacts (for example, EYCP Champions' qualifications and teaching experience, children's prior language and communication skills, child-level exposure to EYCP, FEE status, and EAL status)?

5.6 Are settings able to incorporate EYCP strategies into their practice and sustain them in the longer term?

5.7. What are the perceived and observed impacts of the EYCP on disadvantaged children (i.e., with FEE status)? What are the barriers or facilitators that are specific to those disadvantaged children?

5.8. What are the unintended consequences/impacts of the EYCP programme in intervention nurseries?

5.9. To what extent does the perceived impact in Cohort 1 (2023–2024 school year) differ from the programme impact in Cohort 2 (2024–2025 school year)?

Fidelity, feasibility, and quality assurance of impact evaluation

RQ6: To what extent does the EYCP programme impact evaluation process adhere to the plan?

6.1. Does the nursery setting recruitment process adhere to the recruitment strategy?

6.2. Do PVI, maintained, and SN settings respond similarly or differently to the recruitment strategy?

6.3. Does the children and family recruitment process adhere to the recruitment plan?

6.4. Do baseline and post-test administrators (independent researchers) effectively and appropriately deliver the assessment tasks?

6.5. To what extent does the setting and children recruitment strategy used for Cohort 1 differ from the strategy used for Cohort 2?

RQ7: What is the feasibility of the impact evaluation plan in this efficacy trial and is it feasible to be scaled up?

7.1. Are parents/carers administered language measures acceptable for parents/carers and are they feasible to use in a large-scale trial in the future?

7.2. Are the setting and children recruitment strategies used in the current trial feasible to use in a large-scale trial in the future?

7.3. What challenges/opportunities has the age of the target population (two to three years old) had for delivery of the evaluation?

7.4. What factors may need to be considered in scaling up the intervention to deliver in more nurseries?

Cost evaluation

RQ8: Is the EYCP intervention cost-effective for participating settings?

8.1. To what extent does the EYCP intervention cost for Cohort 1 differ from the cost for Cohort 2?

Good/best practice

RQ9: What are the key features of the EYCP 'good/best practice' settings in the current trial?

Ethics

Ethics approval for the evaluation was initially obtained from the Durham University School of Education Ethics Committee on 3rd March 2023, and an update to use the split cohort design and changes to programme length and sample size obtained ethical approval on 29th August 2023.

Agreement for nursery settings to participate in the trial was obtained through a Memorandum of Understanding (MoU) signed by nursery senior leadership which provided full details about the project, responsibilities, and requirements of the settings. Children were recruited to take part in the trial by the nursery settings sharing an information sheet about the project with parents/carers and talking to them about the project. Parents/carers actively opted for their child's participation in the project by completing and signing a consent form. Recruitment documents including the MoU, information sheets, consent forms, and privacy notices are included in Further appendices.

Registration

This trial was registered on the Open Science Framework (OSF) site in October 2023 (registration number CP9Y3, DOI link **https://osf.io/cp9y3**, with registration amended for protocol v2.0).

Data protection

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

The lawful basis the ET at Durham University relied on for collecting and processing data as part of this project is 'public task', which according to the General Data Protection Regulation (GDPR) Article 6 (1)(e) means 'the processing is necessary for an activity being carried out as part of the University's public task, which is defined as teaching, learning, and research'. This project carried out research and did not process special category personal data.

Privacy notices for setting staff and parents/carers were provided when signing up to the project (see Further appendices). These set out what, and how, data would be collected and processed as part of the project, how it would be stored, who it would be shared with, along with information about participants' ability to withdraw from data processing as part of the project and the retention period for project data. Durham University acted as data controllers for the data collected as part of the evaluation. The East London Research School and Durham University acted as joint controllers of delivery data where it was also used in the evaluation. Oxford University acted as data processors for the Observing Language Pedagogy (OLP) data collected as part of the project. The data retention period for storing and processing personal data is five years beyond the end of the project (April 2030), at which point personal data will be deleted or anonymised by the ET. Anonymous data may be kept indefinitely by the ET and potentially shared with other research teams for future research. At the end of the evaluation study, data was submitted to the EEF's data archive. At this point, the EEF and its archive manager (Fischer Family Trust—FFT Education) became responsible for controlling and processing the evaluation study data.

A Data Protection Impact Assessment (DPIA) was conducted for this project by the principal investigators, describing why the collection of personal data is necessary, the data that will be collected, and the ways that risks will be minimised. A data management plan was put in place to set out the procedures for the ET to follow to ensure data confidentiality throughout the project. A three-way Data Sharing Agreement (DSA): controller to controller was put in place between the DT at ELRS, the ET at Durham University, and each participating nursery to ensure that all teams handled data correctly and that data was shared securely and appropriately. A Data Processing Agreement (DPA): controller to processor was put in place

between the ET at Durham University and Dr Sandra Mathers at Oxford University as data processor for the OLP tool, to ensure that data was handled, processed, and shared securely and appropriately.

Project team

Delivery team (DT)

Fliss James is Director of East London Research School—based at Sheringham Nursery School and Children's Centre. Fliss was joint programme delivery lead and had responsibility for the design, delivery, and management of the project.

Melissa Prendergast is Deputy Headteacher of Sheringham Nursery School and Children's Centre. She strategically leads East London Research School and A Brighter Start: East London's Stronger Practice Hub. Melissa was joint programme delivery lead and had responsibility for the design, delivery, and management of the project.

Meher Faisal is Project Manager at A Brighter Start: East London's Stronger Practice Hub. Meher was responsible for the delivery team's project management.

Rizwana Rahman is the Stronger Practice Hub Administrator at A Brighter Start: East London's Stronger Practice Hub, Sheringham Nursery School and Children's Centre. Rizwana was the key contact for the delivery team during the setting recruitment period.

Eight delivery team mentors also supported the delivery of the programme providing training support and support to settings.

Evaluation team (ET)

Dr Xiaofei Qi is an Associate Professor at the School of Education at Durham University. Xiaofei was the lead principal investigator and had overall responsibility for the evaluation design and protocol development, delivery, reporting, and overseeing the project management.

Vic Menzies is an Assistant Professor at the School of Education at Durham University. She was joint principal investigator and had oversight of the trial design, delivery, and reporting. She was responsible for data management and the cost evaluation.

Dr Rachel Oughton is an Associate Professor in Statistics at the Department of Mathematical Sciences at Durham University. She was a co-investigator and was responsible for the development of the SAP and overseeing the statistical analysis for this trial.

Dr Akansha Singh is an Assistant Professor (Research) in statistics and quantitative methods in the Department of Anthropology. She was a researcher on the project (since September 2024) contributing to the development of the SAP and to the statistical analysis for the trial.

Dr Nadia Siddiqui is a Professor at the School of Education at Durham University. She was a co-investigator and contributing to quality assurance of the evaluation.

Dr Laura Chambers is a Research Assistant at the School of Education at Durham University. She was the project Coordinator (since October 2024) liaising with various stakeholders and had a specific focus on the impact evaluation and data management of the evaluation data.

Dr Dandan Chen was a Research Associate at the School of Education at Durham University. She was the research support officer for the EYCP project with a specific responsibility for the Implementation and Process Evaluation of the trial. She worked on the trial between September 2024 and February 2025.

Rowan Van-Muysen was a Research Assistant at the School of Education at Durham University. He was the project Coordinator liaising with various stakeholders focusing on the impact evaluation and data management of the evaluation

data. He worked on the trial between September 2024 and September 2025. He also supported with IPE data analysis as a casual research assistant.

Dr Yuting Guo was a casual Research Assistant on the project working on data entry and quality assurance of language assessment data, and analysis of IPE data.

Dr Nashwa Ismail was a Lecturer at the School of Education at Durham University. She was the co-investigator until September 2023 and was involved in the planning stage of the trial.

Methods

Trial design

Table 3: Trial design

Trial design, including number of arms		Two-armed cluster randomised waitlist-controlled trial
Unit of randomisation		Nursery
Stratification variable(s) (if applicable)		• Nursery type (Private, Voluntary, and Independent (PVI), maintained and School-based nursery (SN) settings) • Nursery size with 2–3-year-old cohort (Small settings; Large settings[3]) • Geographic location (Local Authorities as proxy for 'mentor area')
Primary outcome	Variable	Receptive vocabulary
	Measure (instrument, scale, source)	• BPVS-III[4] (0–168, raw score), researcher administered, GL Assessment
Secondary outcome(s)	Variable(s)	1) Children's expressive language 2) Children's language and communication skills 3) Practitioners' pedagogical knowledge of language learning and development
	Measure(s) (instrument, scale, source)	1) CDI-III (0–100, raw score, parent/carer administered, Brookes Publishing) 2) WellComm Early Years toolkit (raw score 0–10), researcher administered, GL Assessment 3) OLP tool practitioner self-report (Perceiving: raw score 0–no limit, Naming: raw score 0–8, and Interpretation: raw score 0–8) (Mathers, 2020)
Baseline for primary outcome	Variable	Receptive and expressive vocabulary
	Measure (instrument, scale, source)	• ELIM (0–50, raw score), parent/carer administered, Public Health England • BPVS-III* (0–168, raw score), researcher administered, GL Assessment (BPVS-III will be used only for the sensitivity analysis of the primary outcome by replacing ELIM with BPVS-III)

[3] Small settings are those with fewer children enrolled than the median setting recruited into this trial and large settings are the same or more than the median setting.

[4] At protocol, it was planned to use a combined measure of BPVS-III and CDI-III, and the SAP set out the data conditions required to use this combined outcome (described later). These conditions were not met at analysis stage, and the alternative primary outcome plan described in the SAP was put in place. BPVS-III was therefore used as the primary outcome, and the CDI-III used as a secondary outcome.

Baseline for secondary outcome(s)	Variable	N/A
	Measure (instrument, scale, source)	N/A

*BPVS-III was administered by researchers at baseline with a lower response rate than ELIM due to difficulties of engaging 2-year-olds just starting nursery at baseline; therefore, it was decided not to use it as a baseline outcome in the primary analysis, but instead in the additional (sensitivity) analysis to explore the use of BPVS-III as a baseline outcome in the current study.

The trial used a two-armed cluster randomised waitlist-controlled efficacy trial design with randomisation at the nursery level, to evaluate the impact of the EYCP on progress in the vocabulary, language, and communication skills of children aged 2–3 years in nursery settings (see Table 3). A waitlist-control design was chosen given the EYCP was co-funded with the Department for Education (DfE)'s Stronger Practice Hub (SPH) initiative and consequently there was an interest in all settings accessing the intervention. The SPH initiative was part of the early years education recovery support package to provide advice, share good practice, and offer evidence-based professional development for early years practitioners (DfE, 2023b).

The trial was run over three years, with the intervention being delivered to two different cohorts. The decision to use a split cohort design was made at the setting recruitment phase in 2023 due to not meeting the required number of settings for the full trial to be delivered in one cohort. Cohort 1 settings were recruited and randomised into intervention and control groups in 2023, with intervention settings receiving the EYCP during the 2023/2024 school year and waitlist-control settings receiving it during the 2024/2025 school year. Cohort 2 settings were recruited and randomised into intervention and control groups in 2024, with intervention settings receiving the EYCP during the 2024/2025 school year and waitlist-control settings receiving it during the 2025/2026 school year. Assessment for each cohort took place before the intervention period in September/October and then immediately after the intervention period in June/July.

Participant selection

Nurseries

The trial aimed to recruit 112 nurseries over two years. While the EYCP programme was originally developed to support practitioners in PVI settings, in this efficacy trial, maintained nurseries and school nursery settings in selected local authorities were also included. The reason was twofold:

- to represent the main types of EYFS providers for 2–3-year-olds; and
- to reduce the overall attrition rate which is particularly higher in PVI settings due to reasons such as staff turnover.

As planned in the original Evaluation protocol (Qi *et al.*, 2023), DT led the setting recruitment process, supported by the ET. A detailed setting recruitment strategy was developed by the DT, in line with the EEF-SPH recruitment guidance (see Further appendices for the recruitment strategy template). The trial aimed to recruit 56 nurseries in each cohort (28 in the intervention arm and 28 in the waitlist-control arm), of which at least 12 (21%) were school nurseries or maintained nurseries and 44 (79%) were PVI settings, which would represent the national split of early years providers by type providing 15 hours FEE for 2-year-olds in 2023 (see Education provision: children under 5 years of age; GOV.UK, 2023a).

However, due to setting recruitment challenges faced both in Cohort 1 and Cohort 2, 44 settings were recruited in 2023 (Cohort 1; 23 in intervention group, 21 in waitlist-control group), and 58 settings in 2024 (Cohort 2: 29 settings in each group). In total, 102 settings were recruited (52 in intervention group; 50 in waitlist-control group; 61 PVI (59.8%) settings and 41 school-based and maintained settings (40.2%) for this trial. For each cohort, restricted randomisation was used to allocate the batch of settings for which we had full information at the point of allocation, so that among these settings there was balance across each covariate. The remaining settings were allocated as their information was received, using minimisation, with $p = 0.9$. This accounts for the slight difference in final trial arm sizes.

In the 2023–2024 (Cohort 1) school year, inclusion criteria for settings were that they:

- were in London (specifically Barking and Dagenham, Barnet, Brent, Camden, City of London, Enfield, Greenwich, Hackney, Hammersmith and Fulham, Haringey, Havering, Islington, Kensington and Chelsea, Lambeth, Lewisham, Newham, Redbridge, Southwark, Tower Hamlets, Waltham Forest and Westminster);
- were a PVI, a maintained or a SN nursery offering FEE for 2-year-olds;
- had a minimum of eight children aged two years old at the beginning of the intervention period, who will attend the setting for a minimum of 15 hours a week during the relevant school year;
- were able to commit to releasing one practitioner who works directly with 2–3-year-olds as the EYCP Champion for PD training sessions (including one and a half days face to face and six hours of live online modules), nine hub meetings and a minimum of five support visits from mentors;
- were able to commit to releasing the setting manager or the EYFS lead for three training sessions, in addition to supporting the EYCP Champion throughout the project;
- had not participated in Manor Park Talks or Newham Communication Project; and
- were not taking part in the DfE's Early Years Professional Development Programme or any other Early Years trials funded by the EEF or similar funder.

In the 2024–2025 school year (Cohort 2), some changes were made to the setting recruitment criteria. First, for Cohort 2, the recruitment regions were expanded to achieve the setting recruitment target (Barking and Dagenham, Barnet, Bexley, Brent, Bromley, Camden, City of London, Croydon, Ealing, Enfield, Greenwich, Hackney, Hammersmith and Fulham, Haringey, Harrow, Havering, Hillingdon, Hounslow, Islington, Kensington and Chelsea, Kingston Upon Thames, Lambeth, Lewisham, Merton, Newham, Redbridge, Richmond Upon Thames, Southwark, Sutton, Tower Hamlets, Waltham Forest, Wandsworth, and Westminster). Second, due to the policy changes to childcare support in England (see Childcare choices; GOV.UK, 2023b) whereby all working families with 2-year-olds received 15 hours of FEE from April 2024, the 'offering FEE to 2-year-olds' was removed from the settings recruitment criteria for Cohort 2 settings. Despite the removal of the FEE requirement criteria from the settings in our study, settings were still expected to have places available for 2-year-olds. All the other recruitment criteria remained the same as Cohort 1.

All participating nurseries received the EYCP programme for free: settings assigned to the intervention arm received the programme during the intervention period (Cohort 1: 2023–2024 school year; Cohort 2: 2024–2025 school year), whereas waitlist-control settings received the programme in the following school year (Cohort 1: 2024–2025 school year; Cohort 2: 2025–2026 school year). In addition, there were financial incentives for all participating nurseries in this efficacy trial (i.e., both intervention and waitlist-control settings), with £150 provided at the start of the trial (after baseline assessment) and £250 at the end of the trial (after endline assessment). Moreover, 50% of cover costs were provided for staff attendance at EYCP training in all participating settings.

Nursery practitioners

When signing the setting MoU, participating settings were required to nominate two practitioners (confirmed before baseline assessment) to take part in the EYCP PD and the associated evaluation activities:

- The setting manager/ EYFS lead.
- The EYCP Champion, a room lead/nursery teacher working with 2–3-year-olds.

The nominated EYCP Champions who completed the OLP tool at the end of the EYCP intervention or participated in the IPE interviews/focus groups were provided with a £15 Amazon voucher as recognition for their time for participating in the research.

Children

Child eligibility criteria were that:

- children were aged 2 years old at the start of intervention in Nov 2023 (Cohort 1) or Nov 2024 (Cohort 2);
- children attended nursery for a minimum of 15 hours per week during term time in the 2023–2024 school year (Cohort 1) or the 2024–2025 school year (Cohort 2);
- children were in the class of the nominated EYCP Champion;
- children's parents/carers anticipated that they would remain at the nursery (i.e., they do not foresee they will leave the nursery) for the duration of the trial (until June); and
- children did not have significant Special Educational Needs and Disabilities (SEND) which would prevent them from accessing the assessment and/or would be distressed through completing the assessment.

A change was made to the protocol to include children who didn't have baseline scores on BPVS given the challenges with assessing children at age 2. Children were also included if their parent/carer did not complete the ELIM measure at baseline to maximise final numbers in the trial.

The recruitment of children and families was led by the ET, supported by the DT. When signing the Memorandum of Understanding (MoU) and Data Sharing Agreement (DSA), settings were asked to provide the estimated number of 2-year-old children starting nursery in September as well as an estimated number of children who are eligible for 15-hour FEE at age 2. This allows the ET to have an estimate of the median cohort size of 2–3-year-olds in participating settings.

After returning the MoU and DSA, participating settings were provided with both digital (online via Qualtrics platform) and paper (via postage) information sheets and privacy notices for parents/carers with instructions on opting into the EYCP efficacy trial. In early September, all settings provided the ET with the number of eligible children meeting the first three eligibility criteria and supported the ET in distributing the physical recruitment package including information sheets, privacy notices, and opt-in consent forms to parents/carers of eligible children.

Parents/carers of participating children were also asked to complete two questionnaires, one at the beginning of the intervention in September/October and one at the end of the intervention in June/July; a £15 high street or local grocery store voucher was provided each time for completing the questionnaire as a recognition for participating in the research. The family-level financial incentives for this trial were managed by ET.

In this trial, we were aiming to recruit a minimum of 8 and a maximum of 15 eligible 2–3-year-old children per setting. However, due to recruitment challenges, the minimum number of children per setting was reduced to five to maximise the setting sample size. For Cohort 1 (2023–2024), we recruited 397 children from 44 settings (intervention group: 206 children; waitlist-control group: 191 children). For Cohort 2 (2024–2025), we recruited 505 children from 58 settings (intervention group: 257 children; waitlist-control group: 248 children). In total, we have 902 children with consent forms from parents for this trial over two cohorts (intervention group: 463 children; waitlist-control group: 439 children). (amended for protocol v2.0)

Outcome measures

Baseline measures

At baseline, the BPVS-III (Dunn *et al.*, 2009), a receptive language measure, administered by researchers, together with the parent/carer-reported Early Language Identification Measure (ELIM) (Law *et al.*, 2020; Public Health England, 2020b), an expressive language measure, were used to capture children's receptive and expressive vocabulary attainment before the EYCP intervention. In deviation from the original protocol (Qi *et al.*, 2023), where it was proposed that a combined baseline outcome of the BPVS-III (raw score: 0–168) and ELIM (raw score: 0–50) to 21 would be used to represent children's receptive and expressive vocabulary attainment at baseline, only ELIM raw scores were used as the baseline outcome for the primary analysis. This change was due to the low response rate of baseline BPVS-III scores and difficulties in engaging two-year-olds

with the assessment, especially when new to a nursery setting, which, if combined with ELIM raw scores, would lead to substantial data attrition and a consequent loss of study power (amended for protocol v2.0). There was a significant positive correlation between ELIM and Baseline BPVS (r = 0.45, 95% CI [0.38, 0.51], p < .001), suggesting that children with higher expressive vocabulary at baseline also tended to show higher receptive vocabulary, providing support for the use of ELIM as a baseline measure despite targeting a related but distinct construct.

British Picture Vocabulary Scales—3rd edition (BPVS-III)

The BPVS-III is a standardised vocabulary measure based on the Peabody Picture Vocabulary Test (PPVT) (Dunn and Dunn, 1981). The scale is designed to assess receptive (hearing) vocabulary with children aged 3 to 16 years old and can be used as a baseline measure for children starting nursery or school, as well as a benchmarking and progress-checking assessment. It is used widely by speech and language therapists, educational psychologists, and additional support for learning teachers because vocabulary is an important predictor of other language skills (Dunn *et al.*, 2009). The test involves showing the child a series of images (four on one page) and asking the child to point to the picture that matches a pre-specified word the assessor says to the child. Given that no reading or verbal response from children is needed for the BPVS-III assessment, it is considered suitable for young children and those for whom English is an additional language. It takes a brief time (between five and eight minutes) to complete, enabling children to remain focused during the assessment period. BPVS-III has been standardised in the UK for children aged 3 to 16 years; it has also been used in studies (i.e., ParentChild+; Tracey *et al.*, 2019; 2022) with 2–3-year-olds, and it was regarded as a suitable measure in that there was variation in response and no floor effects were detected. The BPVS-III was administered one to one by an independent researcher who was blind to the intervention condition. Researchers all had prior experience working with young children and received BPVS training from the ET in advance of conducting the assessment. In order to assure quality of assessment delivery, researchers were observed and given feedback in a mock assessment situation following the training and were accompanied by an experienced assessor on their first nursery visit, where they observed assessment delivery before being supervised delivering the assessment and provided with further training or feedback where necessary.

ELIM

As the trial targeted children who were 2 years old at baseline assessment and there were anticipated challenges of child non-engagement and poor concentration with the researcher-administered assessment with children this young, a parent-reported language assessment measure ELIM Word List (at baseline), was also administered to fully capture children's vocabulary attainment at baseline. Previous studies have suggested the feasibility of using parent-reported measures such as ELIM in capturing early language skills in young children (Law *et al.*, 2020; Lindorff *et al.*, 2022; Public Health England, 2020b). The ELIM was developed by Law and colleagues (2020) and has been used previously in England (Public Health England, 2020b). It is suitable for use with the parents/carers of children aged between 2 and 2.5 years. It includes a 50-word list which all children aged 2 to 2.5 would be expected to say (for example, mama) and some which only very few children will say (for example, gentle). Parents indicate the words they have heard their child use. The full list of words can be seen in the Early Language Identification Measure and Intervention: Guidance Handbook published **online**. This ELIM 50-word list was originally distilled from the Communicative Development Inventory (CDI) for use with two-year-olds in Sure Start Programmes in England and was therefore considered an age-appropriate baseline measure in this trial (Law *et al.*, 2023; Roy *et al.*, 2004). An online version of the 50-item vocabulary checklist was created and was hosted on a secure digital platform that parents/carers accessed via a pre-assigned link or QR code. This was completed in September/October before setting randomisation. It is noteworthy that the complete ELIM measure also includes an observation form to record behaviours that the practitioner observed or heard from the child while speaking with his/her parent/carer during the visit, rather than behaviours reported by the parents/carers. However, in this trial, it was decided to use the ELIM Word List without the observation form, given that individual review visits were neither possible nor feasible.

Primary outcome

The primary outcome was planned to be children's receptive and expressive vocabulary attainment, as measured by a combination of the researcher-administered BPVS-III and parent/carer administered MacArthur-Bates Communicative Development Inventory III (CDI-III; Dale *et al.*, 2023) at the end of the EYCP intervention. This would have been the first principal component of the combined BPVS-III and CDI-III data if both of the following conditions were met:

1. The proportion of variance explained by the first principal component was at least 80%.

2. The proportion of complete data for the first principal component was at least 80% (i.e., less than 20% missing for one or both of BPVS-III and CDI-III).

If either of these was not met, the primary outcome was to be BPVS-III, with secondary outcome analysis on CDI-III.

Because more than 20% of the first principal component of the combined BPVS-III and CDI-III data was missing (38.84% missing) and the first principal component explained only 76.68% of the variance, the combined measure failed both pre-specified conditions. Therefore, BPVS-III (which had 18.25% missingness) was used as the primary outcome, and CDI-III was considered as the secondary outcome.

BPVS-III

The BPVS-III was delivered in the same way as at baseline by independent researchers who were blind to the intervention condition.

Secondary outcomes

The trial had three secondary outcomes: 1) children's expressive vocabulary measured using the CDI-III (originally planned to be part of the primary outcome measure), 2) children's holistic language and communication skills measured with the WellComm Early Years screening tool, and 3) practitioners' language-related pedagogical knowledge measured with the Observing Language Pedagogy tool.

CDI-III

As described above, the CDI-III was originally considered as part of the primary outcome measure, but due to the combined measure failing the pre-set criteria, it is used as a secondary outcome. The MacArthur-Bates Communicative Development Inventories (MB-CDIs) (Fenson *et al.*, 2007) are widely used assessment tools for early language and communicative development through parent report. CDI-III (Dale *et al.*, 2023), part of the third edition of the MB-CDIs, is a short tool that measures language and communication skills for children between the ages of 30 and 37 months. The validity of parent-reported child vocabulary measures such as the CDI-III, as well as the potential of such measures to be sensitive enough to detect an effect, has been explored in a similar early years study with an RCT design (Sylva *et al.*, 2021); the authors highly recommended the use of a parent-reported child vocabulary measure alongside a researcher or practitioner assessment of child language (Lindorff *et al.*, 2022). The CDI-III comprises three parts: vocabulary, sentences, and using language; only the 100-item vocabulary checklist section (raw score: 0–100) was used in this trial to measure children's vocabulary as part of the primary outcome. With the author's permission, an online version of the 100-item vocabulary checklist was created for parents/carers to complete in June/July at the end of the EYCP intervention. The CDI-III questionnaire was hosted on a secure digital platform, with parents/carers accessing via a link or QR code. Paper versions were also distributed where this was the preference of the nursery.

WellComm Early Years

The WellComm Early Years toolkit (2010) is a norm-referenced assessment tool that measures the global/holistic language of children from six months to six years old and provides guidance on appropriate support and intervention. It was developed by speech and language therapists in the UK to be used by anyone working with early years children and is both short and simple to administer. The screening tool uses a combination of observation, direct assessment, and carer reports to gain a holistic picture of children's language skills. It consists of ten items related to children's language that correspond to the child's age (with each section covering 6-month age bands). The screening tool uses a traffic light system that bands children by placing them into one of three categories according to their score (RAG rating). Red: Consider referral to a specialist service for further advice/assessment; Amber: Extra support and intervention required; Green: No intervention currently required. The WellComm Early Years Screening tool was administered by independent researchers after the EYCP intervention, following the same recruitment and training as detailed above for the BPVS-III. Each child completed the section relevant to their age at testing with the child's nursery key worker providing responses to the carer report questions.

The WellComm Early Years toolkit raw scores (0–10) for the completed section were used to produce the secondary outcome variable for holistic language and communication skills.

As specified in the logic model (see Figure 1), the EYCP intervention aims to improve children's receptive and expressive vocabulary attainment, as well as language and communication skills through enhanced adult-child interactions and language-rich environments. Despite the primary outcome measure change, the selected child-level outcome measures still broadly align with this model: BPVS-III provides a robust assessment of receptive vocabulary, while CDI-III assesses expressive vocabulary, and WellComm captures holistic language and communication skills. Together, these measures operationalise the key child-level outcomes identified in the programme's theory of change.

Observing Language Pedagogy (OLP) tool

The OLP tool was developed by Dr Sandra Mathers to measure practitioners' procedural knowledge of early years practice which is expected to improve due to the training provided in the EYCP programme as specified in the logic model. Findings from a validation study showed that the OLP reliably measures practitioners' procedural knowledge and that it predicts classroom quality (Mathers, 2020; 2021). A recent study (Mathers *et al.*, 2022) reported that reception class teachers' knowledge, as measured by the OLP tool, predicted gains in children's understanding of vocabulary and sentence structure between ages three and five. The OLP tool, as used in this study, presented practitioners with two videos of real classroom interactions and asked them to note at least five strategies or reflections they noted while watching the videos. The videos provide an authentic context for assessing teachers' pedagogical knowledge accurately and are focused on early language development-related strategies, primarily through adult-child interactions. The tool was designed to capture the facets of perceiving, naming, and interpreting in practitioners. In this trial, the nominated EYCP Champions from all settings (both intervention and waitlist-control settings) were asked to complete the OLP questions online as part of the EYCP Champion endline survey.

The OLP outcome/score has three facets: Perceiving (score: 0–no limit), Naming (score: 0–8), and Interpretation (score: 0–8). Previous research indicates that the OLP data is best represented by a 3-factor/facet model (Perceiving, Naming, and Interpretation) rather than a single dimension; therefore, each facet was used as a dependent variable (Mathers, 2020; Mathers *et al.*, 2022). The OLP tool is not currently commercially available, and it is used in this study with the permission of the author, Dr Sandra Mathers who processed and coded the OLP data and was blind to the intervention group, practitioner details, and setting. A detailed OLP coding strategy and process for this trial was provided by Dr Sandra Mathers (see Appendix F) and can be found in Further appendices.

Sample size

In this trial, recruited settings were split across two cohorts delivered over two years due to recruitment challenges in the first cohort. Sample size calculations were done assuming the full recruited sample across the two cohorts.

Sample size calculations assumed a pre-test/post-test correlation of the primary outcome of 0.60 and a nursery-level intracluster correlation coefficient (ICC) of 0.15. The pre-test/post-test correlation of 0.60 is supported by findings from Parentchild+ and Hanen Learning Language and Loving It (LLLI) studies, which reported similar correlations in language development measures (Tracy *et al.*, 2022; Dimova *et al.*, 2022). The ICC of 0.15 is a standard assumption for cluster randomised trials in educational settings, reflecting the expected variability between nurseries (Robinson-Smith *et al.*, 2023).

At the protocol/design stage, the calculated sample size was 112 settings with an average of 10 children per setting. This allowed for attrition at the setting level (5%) and child level (10%), and would have 80% power to detect an effect size of 0.21 standard deviations between the control and the intervention groups on the primary outcome, and an effect size of 0.34 (2 FEE children per setting)/0.30 (3 FEE children per setting) standard deviations between the control and the intervention FEE subgroups (see Table 14). Previous studies of similar interventions, such as those in the EEF's Early Years Toolkit (n.d.), found effect sizes between 0.53 and 0.61 for communication and language approaches in nurseries. However, many of these studies were not independently evaluated, which could have inflated effect sizes. Independent evaluations such as Nuffield Early Language Intervention (Dimova *et al.*, 2020) or RAND Head Start Impact Study (Puma *et al.*, 2005) tend to yield lower

effect sizes, justifying our target of 0.21, which is equivalent to a moderate effect size. According to the EEF padlock guidance, our trial is powered for an effect size greater than 0.20, designed as a 4-padlock trial (EEF, 2019).

At the randomisation stage, the given sample size (102 settings with a mean of 8.8 children per setting at randomisation) would have 80% power to detect an effect size of 0.22 standard deviations between the control and the intervention groups on the primary outcome. Due to the childcare policy changes in England, the Income Deprivation Affecting Children Index (IDACI) rank (1–3; 30% most deprived) of the child's postcode (collected from their parent/carer) will be used as a proxy for disadvantage for all the sample instead of FEE eligibility which was only available for Cohort 1. For the subgroup of children from a disadvantaged background (as represented by IDACI decile 1–3), the current study would have 80% power to detect an effect size of 0.29 standard deviations between the control and the intervention groups on the primary outcome.

At the analysis stage, the given sample (101 settings with a mean of 7 children per setting) would have 80% power to detect an effect size of 0.26 standard deviations between the control and the intervention groups on the primary outcome. The observed correlation between baseline and endline was 0.48, lower than our assumed value of 0.6. This reduces the power of the trial (and therefore increases the Minimal Detectable Effect Size—MDES). For the subgroup of children from a disadvantaged background (as represented by IDACI decile 1–3), the current study would have 80% power to detect an effect size of 0.38 standard deviations between the control and the intervention groups on the primary outcome (see more details in Table 14).

Randomisation

Nurseries were randomised to intervention or waitlist-control arm after child recruitment and baseline data collection were completed. The minimisation method was used (using the Minirand package in R; Jin *et al.*, 2020), as this is an appropriate way to achieve balance in small trials (Altman and Bland, 2005). The factors controlled were setting type (PVI or school/maintained), number of children recruited (more than 8, 8 or fewer), and local authority (17 levels). For the majority of settings, restricted randomisation was used, selecting an allocation at random from all those possible allocations that had balance across all factors and overall group sizes. As information was received for the remaining settings, they were allocated using minimisation. The probability of allocating to the optimal arm was set to 0.9, to retain some randomness while still achieving balance. This method aims for equal balance across each factor, but does not guarantee an exact 50:50 split. The effect of the different arm sizes (50 vs 52 settings in the initial allocation) on the power is negligible. The randomisation process was carried out by Rachel Oughton, a statistician for the ET who did not have any involvement in the day-to-day delivery of the project.

Statistical analysis

The statistical analysis followed the most recent EEF Statistical Analysis Guidance (EEF, 2022a). All analyses were conducted on an intention-to-treat (ITT) basis, using two-sided significance at the 5% level. ITT analysis was conducted on all children who were randomised, regardless of whether they met the child attendance eligibility criteria (all the children were recruited based on specified eligibility criteria; however, child attendance eligibility was confirmed at endline). A summary is provided below of the planned statistical analysis of primary and secondary outcomes, subgroup analysis, analysis in the presence of non-compliance, and additional analyses. For all the primary and secondary outcomes (BPVS-III, CDI-III, ELIM, and WellComm), raw scores were used for analysis.

Primary analysis

The primary analysis investigated any difference in the primary outcome (endline BPVS-III raw score) at endline and took the form of a linear mixed model at the child level with primary outcome score as the dependent variable.

Baseline raw score (ELIM raw score), group allocation, cohort, and minimisation factors were included as fixed effects in the model. Adjustments for clustering at the setting level were made by including the setting as a random effect. Cohort was included in the analysis, as there were some differences in the way Cohorts 1 and 2 were recruited as well as the way the programme ran for both cohorts, which was then accounted for in the analysis. The minimisation factors included: 1) nursery type (Private, Voluntary, and Independent (PVI), or maintained and School-based nursery (SN) settings); 2) child cohort size

(<median number of children; >=median number of children); and 3) nursery geographic location (the eight EYCP mentors are locally sought/based to support seven settings in a single Local Authority or nearby Local Authorities). Further details are available in the protocol.

Multilevel model equations

$$\text{Post}_{ij} = \beta_{00} + \beta_1\text{Pret}_{ij} + \beta_2\,\text{Group_Allocation}_j + \beta_3\text{Cohort}_j + \beta_4\text{Nursery_type}_j + \beta_5\text{Cohort size}_j + \beta_6\text{*Mentor_group}_j + u_{0j} + e_{ij}$$

Descriptions of the variables and terms

Primary outcome score (Post_{ij}): The dependent variable representing the primary outcome measured at endline for each child.

Pre-test score (Pret_{ij}): β_1 is the fixed effect indicating the effect of the baseline score.

Group_allocation: β_2 is the fixed effect for the group to which the child was allocated (for example, intervention vs control).

Cohort: β_3 is the fixed effect for the cohort to which the child belongs.

Nursery type: β_4 is the fixed effect for the type of nursery (PVI, maintained, or SN settings).

Cohort size: β_5 is the fixed effect representing the size of the child cohort (categorised as less than or greater than/equal to the median number of children).

Mentor group: β_6 is the fixed effect indicating the mentor group.

Random effect (u_{0j}): A random effect accounting for the clustering of children within settings.

Residual error (e_{ij}): The error term representing the unexplained variability at the child level.

This model allowed the investigation of differences in the primary outcome at endline while accounting for the fixed effects of group allocation, cohort, and minimisation factors, as well as the random effect of clustering at the setting level. R software version 4.4.2 was used to conduct this analysis, and the model was fitted using the 'lme4' and 'eefAnalytics' packages.

Secondary analysis

Since the CDI-III raw scores were not part of the primary outcome analysis (as originally planned), this was considered as a secondary outcome, and a separate secondary analysis was conducted considering a similar set of covariates as in the primary outcome analysis.

Secondary analyses therefore investigated any differences in 1) children's expressive vocabulary using the CDI-III raw scores, 2) children's holistic language and communication skills measured with the WellComm Early Years toolkit, and 3) practitioners' language-related pedagogical knowledge measured with the OLP tool. 1) CDI-III and 2) WellComm Early Years toolkit raw scores were analysed using the linear mixed model following the same multilevel model specifications used in the primary analysis. For the CDI-III, ELIM scores at baseline were included in the model as both were measures of expressive vocabulary. Since no baseline data for the WellComm Early Years toolkit was available, only group allocation, cohort, and minimisation factors were included as covariates in the model. Further, baseline ELIM was not included as a baseline covariate in analyses of the WellComm outcome, as WellComm assesses broader language and communication skills. For 3) The OLP outcome, three scores were provided: Perceiving (score: 0–no limit), Naming (score: 0–8), and Interpretation (score: 0–8). Previous research indicates that the OLP data is best represented by a 3-factor/facet model (Perceiving, Naming, and Interpretation) rather than a single dimension; therefore, each facet has been analysed as a dependent variable (Mathers, 2020; Mathers *et al.*, 2022). Given that the OLP outcome is a practitioner measure with one observation per setting, this outcome was analysed using a linear regression model with group allocation, cohort, and minimisation factors as covariates in the model.

No formal statistical corrections were applied for multiple secondary outcomes or subgroup analyses. Such corrections are typically more critical in the context of multiple primary outcomes, where controlling the overall Type I error rate is essential. In this study, secondary outcomes and subgroup analyses were pre-specified in the Statistical Analysis Plan and were conducted mainly to explore potential heterogeneity in intervention effects.

Analysis in the presence of non-compliance

Compliance in this trial is defined as the extent to which the critical ingredients of the EYCP intervention were delivered to and/or received by the target participants (settings/practitioners/children).

Compliance was measured at the nursery level as well as the child level. Each nursery in the intervention arm was assessed for compliance in terms of the key programme inputs/activities as specified in the EYCP logic model in the project evaluation protocol. As agreed with the DT, all core components of the EYCP programme were assumed to be key driving factors and necessary for a successful delivery of the EYCP programme. Table 4 below specifies the different compliance indicators and how they were scored.

Table 4: EYCP compliance criteria

Description	Intervention ingredient/component	Data collection method	Score range
Setting-level compliance indicator			**(0 = no; 1 = Yes; N/A = Not applicable)**
1. Two nominated professionals (setting manager/EYFS lead and nominated EYCP Champion) remain in their EYCP role during the 8-month EYCP intervention.	Core	Collected by the DT	0–1
2. The setting receives a minimum of three in-setting visits from their mentor.	Core	Collected by the DT	0–1
3. The setting undergoes two ITERS-3 audits administered by their mentor.	Core	Collected by the DT	0–1
4. The setting manager/EYFS lead completes required PD training sessions (1.5 days face to face and one live session that is recorded).	Core	Collected by the DT	0–1
5. The setting manager/EYFS lead supports the EYCP Champion and programme implementation in their setting.	Core	Collected by the DT	0–1
6. The EYCP Champion completes required PD training led by the DT (at least 1 day out of 1.5 days face-to-face training sessions).	Core	Collected by the DT	0–1
7. The EYCP Champion completes required PD training led by the DT (at least three out of six live ongoing PD sessions)*.	Core	Collected by the DT	0–1
8. The EYCP Champion attends at least six out of nine hub meetings led by their mentor.	Core	Collected by the DT	0–1
9. The EYCP Champion prepares for and responds to in-setting visits delivered by their mentor.	Core	Collected by the DT	0–1
10. Evidence of the EYCP Champion using EYCP strategies/techniques in their classroom.	Core	Collected by the DT Collected by the ET through practitioner surveys	0–1
Total score (setting core activity)			0–10
Child level compliance indicator		Collected by ET through parents/carers'	**Score range** (1 = 15 to 20 hours; 2 = 21 to 30 hours; 3 = more than 30 hours)

		survey at baseline and endline	
Children's exposure to the EYCP intervention (nursery attendance)**	Core		1–3

* PD sessions were recorded, and practitioners could watch back later, but data was unavailable for those watching pre-recorded sessions; therefore, attendance for at least three out of six live PD sessions has been used as an indicator.
** Nursery attendance pattern was reported by parents/carers both at baseline and endline as a proximal measure for child-level exposure to the EYCP intervention. The endline attendance data was most up-to-date and was used as the child-level compliance indicator. Where this was missing, baseline attendance data was used instead.

Three separate CACE analyses were performed, based on combined setting- and child-level compliance. These analyses used a continuous measure of compliance, minimal and optimal compliance (defined in subsequent paragraphs), to explore the effects of compliance on the primary outcome. Using the combined compliance indicator, a Complier Average Causal Effect (CACE) analysis was performed on the primary outcome to explore the effects of partial compliance. For each CACE analysis, a multilevel model was fitted to the primary outcome including an interaction term between the compliance measure and the group allocation variable along with the other covariates used for the main primary outcome endline BPVS-III analysis. The coefficient of the (group allocation variable) x (compliance indicator) term was used to obtain the required CACE estimate.

Continuous compliance

The setting compliance indicator was represented as a continuous variable in the form of the core components of the EYCP intervention (0–10), as listed in Table 4 above. The child-level compliance indicator was represented as an ordinal variable in the range of one to three. The setting and child compliance variables were rescaled to the interval zero to one (zero was non-compliance and one was full compliance, and there were scores in between). The child and setting level scores were multiplied together to create an individual-level compliance score (which was still between zero and one).

Setting-level compliance

Minimal compliance: A setting achieved minimal compliance if it scored at least one out of the ten core activity items.

Optimal compliance: A setting achieved optimal compliance if it scored at least eight out of the ten core activity items, meeting the following distribution:

- At least two points from items one–three (setting activities).
- At least one point from items four–five (manager activities).
- At least five points from items six–ten (Champion activities).

Child-level compliance

Minimal compliance: Defined as attending at least 15 hours (given that 15 hours is the eligibility criteria).

Optimal compliance: Defined as attending more than 30 hours.

This definition for optimal and minimal compliance at child and setting level is in line with the EYCP logic model.

A combined minimal compliance binary indicator (minimal vs no compliance) was defined for settings that met at least one out of ten core activities and where children attended a minimum of 15 hours.

A combined optimal compliance binary indicator (optimal vs partial or no compliance) was defined for settings that met at least eight out of ten core activities and where children attended more than 30 hours.

Missing data analysis

The amount of missing baseline BPVS-III and ELIM raw scores and primary outcome including BPVS-III raw scores data was summarised. Since the percentage of missing cases for BPVS-III exceeded 5%, the data was checked to see if data was missing in a way that was not correlated with observable variables including baseline attainment, intervention allocation, minimisation factors, and cohort to check if the assumption can be made that data was missing completely at random (MCAR). If data is missing in a way that is correlated with observable variables, then the missing observations are missing at random (MAR). The primary analysis was re-estimated through Multiple Imputation (MI) using Chained Equations (MICE). To do so, the probability that the outcome measure was missing was modelled using a multilevel logistic model that includes the covariates involved in the primary analysis. The covariates thus included baseline ELIM score, intervention allocation, minimisation factors (type of setting, median number of children, and local authority), and cohort (Cohort 1 vs Cohort 2). The significant variables from the baseline ELIM score, intervention allocation, minimisation factors (type of setting, median number of children, and local authority), and cohort (Cohort 1 vs Cohort 2) was then used for an MI process. The minimum number of imputed datasets depends on the fraction of missing information, as suggested by Graham *et al.* (2007). The imputed datasets were used to replicate the main primary outcome analyses (obtained from BPVS-III scores), and the missing data analysis results were compared with the complete data analysis as part of sensitivity analyses.

Subgroup analyses

Subgroup analyses were completed for the following subgroups of children:

1) those with the lowest IDACI deciles (deciles 1 to 3, based on the child's home postcode) (changed from original protocol which specified using entitlement to 2-year-old FEE to define the disadvantaged subgroup, due to government policy change during the project entitling all 2-year-olds to free childcare);

2) those who had English as an additional language (EAL) as indicated on the baseline parent survey that English is not the main language spoken at home; and

3) children who were the lowest (bottom 20%) / best performers (top 20%) in the baseline vocabulary assessment (i.e., below the 20th percentile and above the 80th percentile).

The impact of the programme for these groups was assessed both via the inclusion of the subgroup status and an interaction term between the subgroup status and group allocation in the primary analysis model following similar specifications, and by repeating the primary analysis in each subgroup of children.

All subgroup analyses were conducted for the combined data including both Cohort 1 and 2. The results for the subgroup analysis are provided in the main report while the interaction model results are reported in the Appendices.

Additional analyses and robustness checks

Additional analyses for primary outcome

Difference between Cohort 1 and Cohort 2

The difference between results for Cohort 1 and 2 was assessed by fitting an interaction model including intervention and cohort and checking if there is a significant interaction effect based on *p*-values. As a robustness check, the primary outcome was also analysed separately for each cohort, and these cohort-specific results were compared with those from the primary analysis for the complete sample.

Exclusion based on nursery attendance

Since some children received the intervention and signed up to the evaluation without fulfilling the eligibility criteria of attending nursery for more than 15 hours, an additional sensitivity analysis was conducted for the primary outcome by excluding children who had less than 15 hours nursery attendance to assess the effect of this exclusion on primary outcome results.

Differences by baseline measures

To assess the effect of using different baseline outcomes on primary outcome results, additional sensitivity analysis was conducted replacing the ELIM baseline score with the BPVS-III baseline score and redoing the primary outcome analysis. Conducting a sensitivity analysis by comparing the results using different baseline measures (ELIM and BPVS-III) can help assess the robustness of the findings. By re-running the primary analysis using an alternative baseline measure (BPVS-III instead of ELIM), it can be assessed whether the observed intervention effects depend on the specific baseline measure chosen. If the results remain consistent across both baseline adjustments, this increases confidence that the estimates are stable and not an artefact of the baseline measure used. To interpret any differences in the results, this analysis was done using the BPVS-III baseline score only for the sample for which we have both ELIM and BPVS-III scores. This ensured a thorough comparison and interpretation of the differences, even with varying sample sizes. However, the low data collection rates for BPVS-III may still limit the reliability of this comparison. Attrition in baseline BPVS-III scores could likely be non-random and related to underlying language ability, introducing potential selection bias and limiting the interpretability of sensitivity analyses based on these cases.

Additional analysis for secondary outcome

Additional secondary outcome analysis was conducted using the WellComm assessment RAG rating as a binary outcome variable to look at whether there are fewer children needing additional support as a result of the intervention (i.e., Green = no intervention needed, Amber/Red = additional support is needed and/or consider referral to a Speech and Language Therapist). This took the form of a mixed-effects logistic regression model.

The mixed-effects logistic regression model used is provided below:

$$\text{logit}(P(\text{RAG}_{ij} = 1)) = \beta_{00} + \beta_1 \text{Pret}_{ij} + \beta_2 \text{Group_Allocation}_j + \beta_3 \text{Cohort}_j + \beta_4 \text{Nursery_type}_j + \beta_5 \text{Cohort size}_j + \beta_6 * \text{Mentor_group}_j + u_{0j} + e_{ij}$$

In this coding, the mixed-effects logistic regression estimates the odds of a child being identified as needing additional support. An odds ratio below one for the intervention group would therefore indicate that children receiving the intervention are less likely to fall into the Amber/Red categories and thus less likely to require additional support, whereas an odds ratio above one would indicate the opposite. This binary structure reflects the key practical question of interest™whether the intervention reduces the proportion of children who would be flagged for extra support or referral™making the results directly interpretable for practitioners and stakeholders.

Additional subgroup analysis

Due to no measure of FEE status eligibility in Cohort 2, additional analysis was conducted for the FEE at age two subgroup for Cohort 1 and IDACI 1–3 for Cohort 1. This analysis allowed examination of the variation in the effect size by considering two different sets of disadvantage indicators (FEE vs IDACI).

The effect of the intervention on the 1) subgroup of children who are cared for/taught by staff who have low (<20%)/high (>80%) self-reported confidence in teaching language, and 2) the subgroup of children who are cared for/taught by staff who have higher (level 5 or above) or lower levels of early years qualification (level 4 or below) as reported in the practitioner baseline survey (see IPE design section in protocol), was assessed via the inclusion of the interaction term between the subgroup status and group allocation in the primary analysis model. The response rate for the staff qualifications and confidence was really high; therefore, we have conducted the analysis using an interaction model. Results from this specific analysis were reliable and not affected by the small sample size as initially thought during the analysis stage.

Estimation of effect sizes

Effect sizes based on the difference between the intervention and waitlist-control groups at the post-test are presented as Hedges' *g* with 95% confidence intervals. Following the EEF guidelines (EEF, 2022a), the unconditional standard deviation (SD) in the primary outcome for the pooled sample was used when computing the Hedges' *g* statistic (square root of the sum of the within- and between-cluster variances from the empty linear mixed model will be used to obtain the *SD*). Confidence intervals for the effect size were calculated using the Hedges (2007) formula.

Estimation of ICC

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

The empty multilevel model from which the ICC was estimated is:

$Post_{ij} = \beta_{00} + u_{0j} + e_{ij}$

$Pret_{ij} = \beta_{00} + u_{0j} + e_{ij}$

Where $u_{0j} \sim N(0, \sigma_u^2)$: A random effect accounting for the clustering of children within settings.

And $e_{ij} \sim N(0, \sigma_e^2)$: The error term representing the unexplained variability at the child level.

ICC (ρ) was estimated as

$$\frac{\sigma_u^2}{(\sigma_u^2 + \sigma_e^2)}$$

Implementation and process evaluation

Research methods

The IPE was designed to be descriptive (cross-sectional and longitudinal) and to be delivered alongside, and complement, the impact evaluation. Following EEF guidelines, (EEF, 2022b), the IPE employed different methods to gather data relating to the pre-specified research questions, and the sample size and participants for each element of the IPE study varied from all participating settings (for surveys) to a purposive sample of settings from intervention settings only. Table 5 summarises the data collection methods used to address the IPE research questions.

In Cohort 1 (2023–2024 school year), seven settings were randomly selected to observe the full process of the intervention in settings with different characteristics (for example, setting type, nursery size, and location), making sure that one setting for each paired mentor group (i.e., eight mentors working in four groups) was selected; six of these settings took part in interviews. Another three settings were selected between mid-implementation and end-implementation to form a 'good/best practice' setting subsample, and two settings took part in interviews in Cohort 1.

As agreed with the DT, the identification of 'good/best practice' settings was based on mentors' and the DT's observations of settings' engagement with professional development (PD) training and their implementation of the EYCP programme. These settings were considered to best reflect the EYCP Theory of Change (ToC), demonstrating effective use of available resources and materials to support delivery and progress throughout the intervention. This was irrespective of their starting point, whether they began with strong provision and practice or less developed foundations. The 'good/best practice' subsample was therefore intended to enable an in-depth exploration of the contexts in which settings were 'successfully' implementing the EYCP programme, characterised by particular forms of 'good practice' during the intervention.

In the 2024–2025 school year (Cohort 2), in line with the mentoring arrangement change (i.e., eight mentors, each to support seven settings), 11 settings were randomly selected to observe the full process of the intervention across different setting characteristics (for example, setting type, nursery size, and location): seven settings took part in interviews. Four 'good/best practice' settings were nominated by the DT between mid-implementation and end-implementation; two settings took part in interviews.

Table 5: IPE design, methods of data collection, and analysis overview

Research methods	Data collection methods	Participants data sources (type, number)	Data analysis methods	IPE research question (RQ)	Implementation/ logic model relevance
Longitudinal design (between settings analysis)	Semi-structured focus group	DT team (from ELRS) (*n* = 3) and mentors (*n* = 8). At two timepoints: mid- and end-points of implementation	Combination of inductive and deductive analysis, with analyses grouped thematically according to RQs	RQ1: 1.1–1.4 RQ2: 2.2 RQ5: 5.1–5.2; 5.8–5.9 RQ6: 6.1–6.3; 6.5 RQ7: 7.2–7.3 RQ9	IPE/logic model; Fidelity/ adherence; Context

Cross-sectional design (between settings analysis)	Semi-structured interview and/or focus group	Setting managers/ EYFS leads (Cohort 1 *n* = 3, Cohort 2 *n* = 8)	Combination of inductive and deductive analysis	RQ1: 1.1–1.3 RQ2: 2.1–2.2; 2.4–2.5 RQ3: 3.1–3.2 RQ4: 4.1–4.4 RQ5: 5.3; 5.5–5.8 RQ6: 6.1–6.3 RQ7: 7.3–7.4	IPE/logic model; Fidelity/ adherence; Inputs/outputs; Process; Outcomes; Context; Cost evaluation
Cross-sectional design (between settings analysis)	Semi-structured interview and/or focus group	EYCP Champions (Cohort 1 *n* = 8, Cohort 2 *n* = 8)	Combination of inductive and deductive analysis	RQ1: 1.1–1.3 RQ2: 2.1–2.5 RQ3: 3.1–3.2 RQ4: 4.1–4.4 RQ5: 5.1; 5.3–5.8 RQ6: 6.1–6.3 RQ7: 7.3–7.4	IPE/logic model; Fidelity/ adherence; Inputs/outputs; Process; Outcomes
Cross-sectional design (between settings analysis)	Semi-structured interview and/or focus group	Other relevant practitioners (Cohort 1 *n* = 2, Cohort 2 *n* = 1)	Combination of inductive and deductive analysis	RQ2: 2.1–2.5 RQ3: 3.1–3.2 RQ4: 4.3–4.4 RQ5: 5.4–5.8	IPE/logic model; Fidelity; Inputs/outputs; Outcomes; Context
Cross-sectional design (between settings analysis)	EYCP training session observations	DT with EYCP Champions and/or setting managers/EYFS leads from intervention settings 1.5 day observations of in-person PD training 6 1-hour observations of online PD training 26 1-hour hub meeting observations	Descriptive/thematic analysis; Content analysis	RQ1: 1.1–1.5	IPE/Logic model; Fidelity/adherence; Input/output
Longitudinal design (between settings analysis)	Administrative records of training attendance and engagement	All intervention settings (*n* = 51)	Descriptive analysis; Inferential analysis	RQ1: 1.1; 1.5 RQ7: 7.3–7.4	IPE/Logic model; Compliance; Context
Longitudinal design (between settings analysis)	Nursery learning environment observation (ITERS-3)	All intervention settings (*n* = 51)	Descriptive analysis	RQ5: 5.2; 5.9	IPE/Logic model; Fidelity/adherence; Input/output; Context; Outcomes
Longitudinal design (between settings analysis)	Baseline usual practice surveys	Control and intervention nurseries (Cohort 1 = 34, Cohort 2 = 49)	Descriptive analysis; Inferential analysis Content analysis; Thematic analysis	RQ3: 3.1–3.2 RQ4: 4.1–4.4	IPE/Logic model; Usual practice; Context; Compliance

Longitudinal design (between settings analysis)	Baseline and endpoint practitioners' confidence surveys	All control and intervention nurseries (baseline *n* = 75; endpoint *n* = 54)	Descriptive analysis; Inferential analysis	RQ5: 5.1; 5.3; 5.5; 5.8	IPE/Logic model; Outcomes; Context
Longitudinal design	Cost evaluation workshops	DT team (*n* = 2)	Cost table for ingredients of intervention	RQ8: 8.1	Cost evaluation
Cross-sectional design	Endpoint nursery cost evaluation survey	Intervention setting managers/EYFS leads (*n* = 29) EYCP Champions (*n* = 32)	Descriptive analysis	RQ8: 8.1	Cost evaluation
Longitudinal design	Semi-structured focus group	Language assessors from ET (*n* = 26)	Thematic analysis	RQ6: 6.1–6.5 RQ7: 7.1–7.4	Feasibility and quality assurance
Cross-sectional design (between settings analysis)	Observation of child assessment sessions	ET assessors for child assessment (*n* = 38)	Descriptive analysis; Content analysis*	RQ6: 6.4 RQ7: 7.3	Feasibility and quality assurance
Good/best practice case study (within settings analysis)	Utilising all data collection methods listed above	DT nominated settings between mid- and end-implementation (Cohort 1 *n* = 2, Cohort 2 *n* = 2)	Thematic analysis; Grounded theory approach	RQ9	IPE/Logic model; Context; Good practice

*Note: There was a deviation from the study protocol in that inter-rater reliability (IRR) analysis using percentage agreement and Cohen's kappa had initially been proposed for quality assurance purposes. However, in practice, quality assurance of language assessors was achieved through a two-way approach comprising assessor training and the in-setting assessment quality assurance process. Observations of assessors during in-setting assessment practice were conducted for all assessors on the first day of assessment delivery using a standard Quality Assurance tool (see Further appendices). The QA process identified any issues in assessment delivery and enabled in situ support and corrective guidance to be provided. Where the observer remained dissatisfied with the assessor's delivery following a morning of in situ training and support, the assessors did not proceed to deliver further assessments.

Baseline and endline surveys for practitioners

Baseline and endline surveys completed by setting staff captured the following measures:

- Confidence and skills in supporting early language development (nominated EYCP Champion; baseline and endline).
- General practice and knowledge of specific EYCP strategies and approaches (nominated EYCP Champion; baseline only).
- Usual continuing professional development (CPD) practices within the setting (setting manager/EYFS lead; baseline and endline).
- Practitioner background characteristics (setting manager/EYFS lead and nominated EYCP Champion; baseline and endline).

- Evaluation of the EYCP programme (intervention settings only; endline).

As the EYCP ToC expected practitioner skill and knowledge in supporting children's language development to improve as part of the programme, it was important to measure this before and after the intervention. To this end, an adapted version of the Practitioner Confidence and Skills in Supporting Children's Language scale (Lindorff *et al.*, 2022) was used. This measure contained two sections. The first section captured practitioners' confidence in knowledge and skill, and asked participants to answer 12 items relating to supporting children's language development. These items used a 5-point Likert scale ranging from 'not at all (0)' to 'very much (4)' in terms of how confident the practitioner feels about each item.

The second section contained items on the general practice of how practitioners interact with individual children in supporting language and communication skills. This was measured through nine items related to practice when interacting with children, looking at how frequently each activity is done using a 5-point Likert scale from 'never (0)' to 'very often (4)' for each item (hereafter referred to as the *General Practice* scale).

The reliability analysis was conducted on the multi-item scales assessing practitioners' confidence and general practice (see Table 6). The 12-item Confidence scale showed excellent internal consistency at both baseline and endline, with Cronbach's alpha (α) values of .936 (Cohort 1), .888 (Cohort 2), and .907 (combined) at baseline, and .927 (Cohort 1), .863 (Cohort 2), and .890 (combined) at endline. The 9-item General Practice scale demonstrated acceptable internal consistency at baseline (α = .919 for Cohort 1, .879 for Cohort 2, .896 combined). The surveys also contained quantitative and qualitative questions regarding practitioners' usual practice and CPD to investigate the context of the delivery of the programme. This allowed the study to look at how EYCP is different to usual practice and the nursery activity that EYCP will be replacing. The baseline practitioner survey also included items asking about total years of experience in Early Years, highest qualification, type of qualification, and role in the nursery. The baseline and endline surveys can be found in the Further appendices.

Table 6: Reliability tests on combined confidence (baseline and endline), and general practice (baseline)

Survey Scale	Stage	*N* of Items	Cohort 1 (α*)	Cohort 2 (α)	Combined (α)
Confidence	Baseline	12	.936	.888	.907
	Endline	12	.927	.863	.890
General Practice	Baseline (only)	9	.919	.879	.896

*Cronbach's alpha (α) indicates internal consistency reliability.

Table 7 and Table 8 present survey response rates from practitioners in the confidence measure and general practice measure respectively. Response rates were higher at baseline than at endline across all groups, with the highest engagement observed in intervention settings at both timepoints. There was a notable decrease in practitioners' responses to the confidence measure between baseline and endline.

Table 7: Confidence in supporting language survey responses from practitioners (EYCP Champions and nominated Champions)

Cohort	*N* settings	Baseline Responses (*N*)	Baseline Responses (%)	Endline Responses (*N*)	Endline Responses (%)
Cohort 1	43	34	79	20	47
Cohort 2	58	51	88	35	60
PVI*	57	48	84	32	56
SN**	44	37	84	23	52

Control	50	38	76	20	40
Intervention	51	47	92	35	69
Total	**101**	**85**	**84**	**55**	**54**

*PVI: Private, Voluntary, and Independent (PVI) settings; ** maintained and School-based nursery (SN) settings

Table 8: Usual practice in supporting language survey responses from practitioners (EYCP Champions and nominated Champions)

	***N* settings**	**Baseline Responses (*N*)**	**Baseline Responses (%)**
Cohort 1	43	34	79
Cohort 2	58	49	84
PVI*	57	46	81
SN**	44	37	84
Control	50	38	76
Intervention	51	45	88
Total	**101**	**83**	**82**

*PVI: Private, Voluntary, and Independent (PVI) settings; ** maintained and School-based nursery (SN) settings

Administrative records (all intervention settings)

Administrative records (provided by the DT) of practitioner attendance at PD training and support sessions were used to measure the extent to which settings complied with the PD training and engaged with PD support. Administrative records from EYCP mentors from their mentoring sessions with each setting were used to capture fidelity to the delivery of the programme's techniques and progress through the nine hub meetings and five nursery visits. These were used to triangulate findings from other IPE data collection methods.

The DT kept records of manager and Champion attendance across all strands of the EYCP including full-day and half-day in-person training events, online modules, hub meetings, and mentor visits. See below for the breakdown of each training component:

- Face-to-face training events: two (one full day at the start of training, one half day at the midpoint).
- Online training modules: six.
- Hub meetings: nine (alternating in person (five) and online (four)).
- Mentor visits to setting: five (two to conduct ITERS evaluation at baseline and endline, three support visits throughout the training).

Attendance was calculated for managers and Champions for each setting. Managers were expected to attend the two in-person training sessions and one online module. Attendance for managers was reported for each session (each scored 0/1) and a total attendance score (maximum score of 3).

Champions were expected to attend all sessions and host all support visits. Attendance was recorded as follows:

- Two in-person training sessions (each scored 0/1).
- Total attendance at online modules (maximum = 6).

- Total attendance at hub meetings (maximum = 9).
- Hosting mentor visits (maximum = 5).
- Overall attendance score (maximum = 22).

Observation/audit of nursery learning environment (all intervention settings)

As part of the EYCP programme, ITERS-3 (Infant/Toddler Environment Rating Scale) observations were conducted by the Delivery Team (DT; mentors) at both the beginning and end of the programme across all intervention settings. These observations assessed the quality of the learning environment, with a particular focus on supporting the development of a language-rich environment, and were used to provide formative feedback to settings on their progress. The resulting data was shared with the ET to enable triangulation with impact evaluation outcomes and self-reported data collected through the IPE. In total, 51 settings completed the ITERS-3 assessment at baseline and 49 at endline (see Table 9 for a detailed breakdown).

Table 9: ITERS-3 (Infant/Toddler Environment Rating Scale) observations by mentors

Cohort	Baseline settings (*N*)	Endline settings(*N*)	PVI* settings (*N*)	SN** settings (*N*)
Cohort 1	22	21	7	15
Cohort 2	29	28	22	7
Total	51	49	29	22

*PVI: Private, Voluntary, and Independent (PVI) settings; ** maintained and School-based nursery (SN) settings

Observations of EYCP training

Two half-day observations of the face-to-face EYCP training (on induction day) were conducted by two researchers from the ET, as part of quality assurance (QA) procedures. These QA observations aimed to assess consistency in programme delivery and to support alignment between observers. Both researchers independently applied a structured observation framework, and their assessments were compared to identify any discrepancies. Differences in ratings were discussed and resolved through consensus, supporting a shared understanding and consistent application of the assessment criteria.

Using a semi-structured observation schedule (see Further appendices), observations of EYCP training sessions were conducted to assess whether professional development (PD) training was delivered as manualised and the extent to which sessions adhered to the pre-specified delivery plan. In addition, cross-sectional observations of mentor-led hub meetings were conducted by researchers from the ET using the same semi-structured observation schedule. A total of eight hub meetings were observed between January and June for Cohort 1, and 20 for Cohort 2. These observations examined whether the EYCP mentoring support was delivered as intended, the degree of adherence to the pre-specified session plans, and the consistency of delivery across mentors. Observations covered a subset of sessions across all training components, enabling assessment of fidelity and consistency of delivery across cohorts and delivery modes (see Table 10).

Table 10: Observations of EYCP PD training and hub meetings across both cohorts

Cohort	Training activity	Maximum number of sessions	Number observed
Cohort 1	In-person training	2 (one full day at the start of the programme and one half day midway through)	2
	Online training modules	6	4 (modules 1, 2, 3, and 6)
	Hub meetings	36 (9 meetings per 4 hubs)	8 (meetings 3, 4, and 9)

Cohort 2	In-person training	2 (one full day at the start of the programme and one half day midway through)	2
	Online training modules	6	3 (modules 3, 4 and 6)
	Hub meetings	63 (9 meetings per 7 hubs)	20 (meetings 1, 4, 6, 7, 8, and 9)

Interviews and/or focus groups

Semi-structured interviews or focus groups with practitioners (setting managers/EYFS lead, EYCP Champions, and a third practitioner when applicable; all settings approached to take part in interviews were invited to include other practitioners who were being trained by the EYCP Champion via the informal 'buddy system' whereby training was cascaded down to the wider setting workforce) from intervention settings were undertaken cross-sectionally between mid-implementation and end-implementation of the EYCP (online via a video conferencing software). Practitioners from the randomly selected and best/good practice settings as described above were invited to interview (see Table 11 for full details of those interviewed). These interviews investigated how the EYCP training was delivered, how the setting had implemented the EYCP programme through the year, the fidelity of delivery, adherence to the implementation plan, any difficulties with implementing the programme, and gathered practitioner views of the acceptability and impact of the programme.

Semi-structured focus groups or interviews with the DT (including the ELRS team and the mentor team) captured their views of the EYCP programme delivery across the project, any challenges they faced, any changes they made to the programme during delivery, and how actual delivery compared to their expectations. Focus groups and interviews with the DT were undertaken longitudinally at two timepoints for each cohort: the first in February/March at the mid-implementation stage, and second at the end of the intervention in July for mentors and in September for the core ELRS team. Mentor interviews took place individually or in groups of two or three. All eight mentors were interviewed at the midpoint for Cohort 1 and six were interviewed at the second timepoint, while for Cohort 2 all eight mentors were interviewed at both timepoints. All interviews with mentors were conducted online using video conferencing software.

Semi-structured focus groups with the ET were conducted at the end of the evaluation to capture the impact evaluation adherence to the evaluation plan and the quality assurance process after the intervention was delivered in August/September of both cohorts (2024 for Cohort 1 and 2025 for Cohort 2).

All interviews and focus groups were audio-recorded for transcription and quality assurance purposes. Interviews and focus groups were conducted by two researchers either together or separately, and a third researcher observed two live interview/focus group sessions per researcher and monitored the interviews/focus group sessions throughout the process for quality assurance purposes. The third researcher sampled 20% of interview/focus group transcripts and checked against the recording to identify mistakes, such as missed and misheard words, to provide an independent interpretation of the information content.

Table 11: Summary information about interview participants

Cohort	Setting ID	Setting type	Randomly Selected (RS) or 'Best Practice'(BP)	Manager/EY Lead (Y=Yes/N=No)	EYCP Champion (Y=Yes/N=No)	Practitioner (the Buddy) (Y=Yes/N=No)
Cohort 1	N09	SN	BP	Y	Y	N
	N11	PVI	RS	N	Y	N
	N15	SN	RS	N	Y	N
	N17	SN	BP	Y	Y	N

	N25	SN	RS	Y	Y	N
	N35	SN	RS	N	Y	N
	N37	SN	RS	N	Y	Y
	N44	SN	RS	N	Y	N
Cohort 2	S08	PVI	RS	Y	Y	N
	S17	SN	RS	Y	Y	Y
	S17	SN	BP	Y	Y	Y
	S33	PVI	RS	Y	Y	N
	S35	PVI	RS	Y	Y	N
	S39	SN	RS	Y	Y	N
	S46	SN	RS	Y	Y	N
	S51	PVI	RS	Y	Y	N
	S78	PVI	BP	Y	Y	N
Total	16 settings*	6 PVI** settings 10 Maintained and SN*** settings	RS = 13 BP = 4	12 managers/EY leads = 12	16 EYCP Champions	3 practitioners

* Setting S17 were interviewed twice (once as random and again as good practice); **PVI: Private, Voluntary, and Independent (PVI) settings; *** maintained and School-based nursery (SN) settings

Analysis

Interviews and/or focus group data

All data collected in interviews and focus groups was analysed using thematic and content analysis and was initially coded deductively based on the project's research questions and then inductively looking for any unexpected emerging themes. We followed the stages outlined by Braun and Clarke (2006): detailed familiarisation; generating initial codes; searching for themes; reviewing themes; defining and naming themes; and data reporting. The original transcripts were coded deductively to develop an initial coding framework based upon the specified IPE dimensions and research questions, and the development and refinement of the coding framework was iterative throughout analysis.

Survey data

Descriptive statistics are presented for the quantitative survey questions showing the exploration of data by randomised group, setting type, practitioner role, and practitioner experience as well as by timepoint (where relevant). Continuous measures are reported as a mean and standard deviation, while categorical data is reported as a count and percentage. Qualitative survey responses were coded deductively using thematic analysis in line with the research questions. Response rates to the surveys are included in Table 12 below.

Data from the 'Practitioner Confidence and Skills in Supporting Children's Language' measure (as specified in the baseline and endline surveys for practitioners section) was compared between intervention conditions (intervention vs control), and between pre- and post- intervention status; data from the 'general practice scale' at baseline was compared between intervention and control conditions; Champions' 'early language development confidence and knowledge' outcome data was initially intended to be analysed as a dependent variable using a linear mixed-effects model. This model would have adjusted for nursery-level minimisation factors (nursery type, child cohort size, and location), as well as cohort and highest qualification of the respondent as fixed effects, with nursery included as a random effect. However, due to the smaller-than-

expected sample size, this analysis was not undertaken. Instead, inferential statistical analyses were conducted to examine differences between groups and over time.

Table 12: Complete survey responses from managers/EY leads and practitioners (EYCP Champions and nominated Champions)

Respondent		*N* settings	*N* Responses Baseline	% Responses Baseline	*N* Responses Endline	% Responses Endline
Manager/EY lead	Cohort 1	43	31	72	19	44
	Cohort 2	58	50	86	41	71
	PVI	57	45	79	33	58
	Maintained	44	36	82	27	61
	Control	50	36	72	31	62
	Intervention	51	45	88	29	57
	Total	**101**	**81**	**80**	**60**	**59**
EYCP Champions or nominated Champions	Cohort 1	43	34	79	19	44
	Cohort 2	58	51	88	34	59
	PVI	57	48	84	30	53
	Maintained	44	37	84	23	52
	Control	50	38	76	20	40
	Intervention	51	47	92	33	65
	Total	**101**	**85**	**84**	**53**	**52**

Observation data

Observation data from EYCP training and mentoring sessions was analysed deductively using thematic and content analysis to look at the level of consistency in the delivery of training sessions between mentors and the engagement of participants.

ITERS-3 audit/observation data

Descriptive statistics are presented for all intervention settings to explore the differences between two timepoints: the beginning of the intervention and the end of the intervention. Inferential statistical analyses were conducted to examine differences between groups (intervention vs control; Cohort 1 vs Cohort 2) and over time (baseline vs endline). As ITERS-3 was administered by the DT (mentors), the independence of the data and findings should be taken into careful consideration.

Results from all IPE data collection methods (observations, interviews/focus groups, surveys, administrative data) have been triangulated and synthesised from the themes and presented as answers to the IPE research questions under the main IPE dimensions.

Participant coding procedure

The names of setting staff (Champions, managers, and other practitioners), mentors, and assessors from the ET were anonymised and replaced with unique ID codes. The coding procedure was as follows.

Champion codes were made up of a C (Champion), followed by the letter N or S (N = Cohort 1, S = Cohort 2), and a number from 01–78 which corresponds to their setting ID. For example, CN17. If a setting had more than one Champion (for example,

a change in role due to staff sickness), a letter A or B was applied to differentiate between the two (for example, CS54A, CS54B). Manager codes follow the same procedure as Champions but begin with an M for example, MS33. Mentor codes are M followed by a number between 1 and 8 (for example, M3). Assessor codes are made up of the letter A (assessor) followed by the number 1 (Cohort 1) or 2 (Cohort 2), then B or E (baseline or endline), and then a number from 01 to 18 (for example, A1E15). The different hub groups were also given unique ID codes beginning C01/C02 (Cohort 1 or 2) followed by the letter H (hub) then a number between 01 and 07 (for example, C01H2).

Cost evaluation

Costs

Data regarding the delivery costs for the EYCP programme was collected in two cost workshops with the delivery and evaluation teams. The workshops explored the cost ingredients of delivering the programme, assumptions for delivering the programme as a three-year model, and how the delivery model for evaluation was different to what it would be delivered outside of a trial. The first semi-structured workshop explored delivery costs for Cohort 1 delivery and was conducted online in May 2024. The second semi-structured workshop exploring changes in delivery costs for Cohort 2 was conducted in person in September 2025 after completion of Cohort 2 delivery. The delivery team shared updated cost spreadsheets with the evaluation team following the workshops which established costs for delivering the programme and providing backfill cover for nurseries to engage with the programme for both cohorts.

Costs associated with time and resources required for setting staff for training and ongoing programme support were explored with the delivery team at the beginning of the programme. The cost evaluation used expected time requirements for the mandatory elements of the programme including attendance at training sessions, hub meetings, and mentor visits, and preparation and planning time for implementation. Backfill costs were provided by the delivery team to nurseries for this amount of time. Survey questions to senior leader/managers and EYCP Champions also explored reported time for preparation for delivery of EYCP in the setting, for meetings between Champions and senior leaders/managers and for engagement of senior leaders/managers in non-mandatory activities. Only 23/87 EYCP Champions provided a specific weekly amount of time for preparation and planning to deliver the EYCP programme. The average of this time (excluding an impossible outlier) has been used in estimates of the weekly planning time required; however, this may be an over-estimate if the majority of survey respondents did not answer the question due to not needing additional planning time. Surveys and interviews with EYCP Champions also explored any other costs settings had in implementing the EYCP programme.

The cost per child per year was established based on three delivery models:

- Cohort 1 delivery model with mentors paired and delivery to 22 settings.
- Cohort 2 delivery model with mentors working individually and delivery to 49 settings.
- Ongoing delivery model beyond the trial with mentors working individually and delivery to 56 settings (mentors supporting 7 settings each).

To determine the cost per child per year, the delivery team costs for providing the EYCP programme were established, as well as the cover costs provided for settings to engage with the programme for the first year of delivery as in the project. A cost for delivery of the programme over the second and third year was estimated to include continued delivery of the programme in the setting without covering costs, no training, or resource requirements but the provision of three one-hour EYCP alumni network peer support meetings (optional for settings) led by the delivery team leads. For each delivery model, the total costs for the delivery across three years for each participating setting were calculated. Then the average cost for each year per setting was calculated. This was then divided by the average number of 2-year-olds expected in the settings across both cohorts as reported in the MoU at sign-up. Sensitivity analysis was conducted to explore the cost for the minimum number of two-year-olds reported across the two cohorts and the maximum number of two-year-olds reported across both cohorts.

Timeline

Table 13: Timeline

Dates	Activity	Staff responsible/ leading
Nov 22–Jan 23	Trial set-up	ET, DT, and EEF
Feb–Sept 23	Protocol and ethics	ET
EYCP Trial in Cohort 1: 2023–2024 School Year		
Feb–Sept 23	Recruitment of settings	DT (with ET support)
Sept–Oct 23	Recruitment of settings and parent/carers	DT (with ET support)
July–Sept 23	Recruit and train assessors for child language assessment at baseline	ET
Sept–Nov 23	Baseline surveys for nursery practitioners	ET
Sept–Nov 23	Pre-test for children	ET
Oct/Nov 23	Randomisation	ET
Nov 23–Jun 24	Delivery of the EYCP	DT
Nov–Dec 23	IPE Observations I	ET
Jan–April 24	IPE Observations II; practitioners interview/focus group	ET
Feb/March 24	IPE developer interview/focus group with DT (ELRS team and mentors)	ET and DT
May–July 24	IPE Observations III; practitioners interview/focus group	ET
May–Jun 24	Recruit and train assessors for child language assessment at endline	ET
Jun–Jul 24	Post-test for children; endline surveys for nursery practitioners	ET
Jul 24	IPE developer interview/focus group with DT (mentors)	ET and DT
Sept 24	IPE developer interview with DT (ELRS team)	ET and DT
EYCP in Cohort 2: 2024–2025 School Year		
Oct 23–Sep 24	Recruitment of settings	DT (with ET support)
Sept–Oct 24	Recruitment of settings and parent/carers	DT (with ET support)
July–Sept 24	Recruit and train assessors for child language assessment at baseline	ET
Sept–Oct 24	Baseline surveys for nursery practitioners	ET
Sept–Oct 24	Pre-test for children	ET

Oct/Nov 24	Randomisation	ET
Nov 24–Jun 25	Delivery of the EYCP	DT
Nov/Dec 24	IPE developer interview with DT	ET and DT
Nov–Dec 24	IPE Observations I	ET
Jan–April 25	IPE Observations II; practitioners interview/focus group	ET
Feb–Mar 25	IPE developer interview/focus group with DT (ELRS team and mentors)	ET and DT
May–July 25	IPE Observations III; practitioners interview/focus group	ET
May–Jun 25	Recruit and train assessors for child language assessment at endline	ET
Jun–Jul 25	Post-test for children; endline surveys for nursery practitioners	ET
Jul 25	IPE developer interview with DT (mentors)	ET and DT
Sept 25	IPE developer interview and cost workshop with DT (ELRS team)	ET and DT
Sept 25–Jul 26	Delivery of intervention to Cohort 2 waitlist settings	DT
Analysis and reporting		
Sept 25–Jan 26	Data analysis, report writing	ET
Feb 26	Submit evaluation report	ET
October 26	Publication of evaluation report	EEF
October 26	Data Archiving	ET

Impact evaluation results

Participant flow including losses and exclusions

In 2023/2024 (Cohort 1), 4309 settings were contacted and invited to take part in the study. Of these, 140 expressed interest in the study; Memoranda of Understanding (MoU) were sent to 66 settings after an initial phone conversation to check setting eligibility criteria[5]. Following that, 59 eligible settings signed and returned MoUs. Of the 59 settings with signed MoUs, 15 settings withdrew prior to randomisation (10 settings were not responding to children recruitment; 5 settings withdrew due to setting readiness issues such as change of management team and staffing issues). So 44 settings were randomised into the study (21 control and 23 intervention). One intervention setting withdrew prior to the start of training and rejoined the study and re-randomised in Cohort 2. In 2024/2025 (Cohort 2), 13,056 settings were contacted and invited to take part in the study in 2024/2025 (Cohort 2). Following that, 179 settings expressed interest in the study and MoUs were sent out to 89. So 80 settings signed and returned an MoU; 22 settings withdrew prior to randomisation (10 settings were not responding to children recruitment; 7 settings did not meet the minimum number of children recruitment; 5 settings were unable to take part due to staff sickness leave, maternity leave, and staff shortage issues). So 58 settings were randomised into the study (29 control and 29 intervention). One intervention setting withdrew in Feb 2025 during training due to staffing issues.

[5] Settings with expressions of interest (EOIs) that were not reachable for eligibility checks or did not meet the criteria were excluded at this stage. The primary reasons for exclusion were location (area) and the absence of provision for two-year-olds.

Figure 2: Participant flow diagram (2 arms)

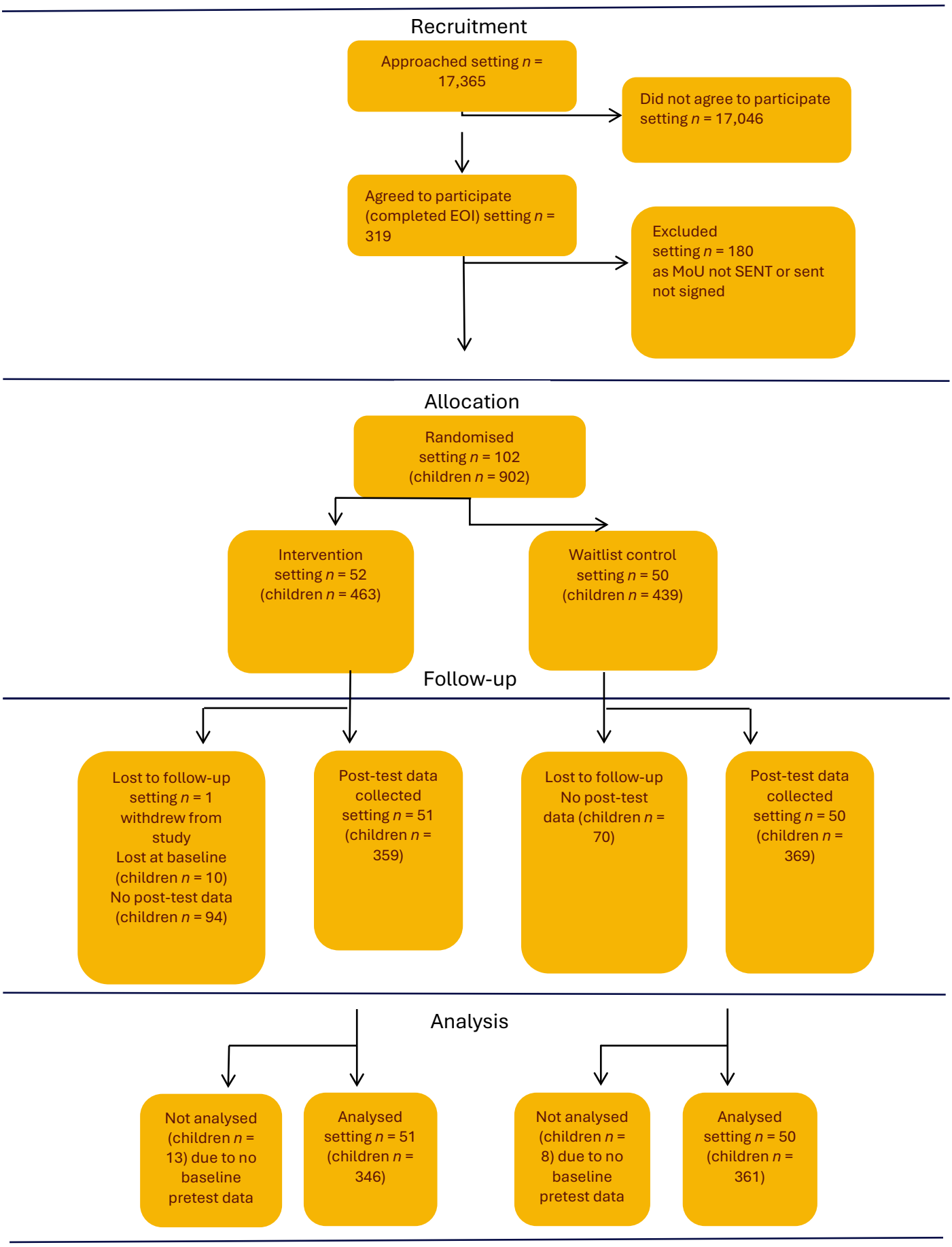


Table 14: Minimum detectable effect size at different stages

		Protocol		Randomisation		Analysis	
		Overall	**FEE**	**Overall**	**IDACI decile (1–3)**	**Overall**	**IDACI decile (1–3)**
Minimum Detectable Effect Size (MDES)*		0.21	0.34/0.30	0.25*	0.41	0.26	0.38
Pre-test/ post-test correlations	level 1 (child)	0.60	0.60	0.60	0.60	0.48	0.44
	level 2 (setting)	N/A	N/A	N/A	N/A	N/A	N/A
Intracluster correlation coefficient (ICCs)	level 1 (child)	N/A	N/A	N/A	N/A	N/A	N/A
	level 2 (setting)	0.15	0.15	0.15	0.15	0.15	0.21
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		9	2/3	8.8	3.4	7	3.47
Number of settings	Intervention	53	53	52	52	51	36
	Waitlist-control	53	53	50	50	50	40
	Total	106	106	102	102	101	76
Number of children	Intervention	477	106/159	463	194	346	108
	Control	477	106/159	439	157	361	156
	Total	954	212/318	902	351	707	264

* Estimates at protocol and randomisation stages assume 5% setting-level attrition and 10% child-level attrition at post-test.

At protocol/design stage, the calculated sample size was 112 settings with an average of 10 children per setting. This allowed for attrition at the setting level (5%) and child level (10%), and would have 80% power to detect an effect size of 0.21 standard deviations between the control and the intervention groups on the primary outcome, and an effect size of 0.34 (2 FEE children per setting)/0.30 (3 FEE children per setting) standard deviations between the control and the intervention FEE subgroups (see Table 14).

At randomisation stage, the given sample size (102 settings with a mean of 8.8 children per setting at randomisation) would have 80% power to detect an effect size of 0.25 standard deviations between the control and the intervention groups on the primary outcome. Due to the childcare policy changes in England, the Income Deprivation Affecting Children Index (IDACI) rank (1–3; 30% most deprived) of the child's postcode (collected from their parent/carer) was to be used as a proxy for disadvantage for all the sample instead of FEE eligibility which was only available for Cohort 1 (see Table 14). At randomisation stage, for the subgroup of children from a disadvantaged background (as represented by IDACI decile 1–3, with 3.4 children per setting), the study would have 80% power to detect an effect size of 0.38 standard deviations between the control and the intervention groups on the primary outcome.

At the analysis stage, the MDES for the whole sample was found to be 0.26, which was larger than the MDES estimated at both the protocol stage (0.21) and the randomisation stage (0.25). This increase in the MDES was largely attributable to higher-than-estimated attrition rates at child level (21.6%), as well as a lower-than-expected pre–post correlation for the primary outcome measure (level 1 child correlation: BPVS/BPVS at protocol/randomisation = 0.60; ELIM/BPVS at analysis = 0.48). For the subgroup of children from a disadvantaged background (as represented by IDACI decile 1–3), the current study would have 80% power to detect an effect size of 0.41 standard deviations between the control and the intervention groups on the primary outcome, with a sample of 264 children from 76 settings (3.7 children per setting). It is noteworthy that the number of settings included in the IDACI subgroup analysis differed between the randomisation and analysis stages, as not all settings included children within the IDACI 1–3 group.

A key consequence for interpretation is that the IDACI subgroup analysis may not fully reflect the originally randomised sample of settings. Because only settings with children in the IDACI 1–3 group were included in the analysis, the subgroup comparison was based on a reduced and potentially different subset of settings. This may have reduced statistical power and precision, and could limit the generalisability of findings to all randomised settings. In addition, if the excluded settings differed systematically from those included, there is potential for selection bias, meaning subgroup results should be interpreted cautiously and considered exploratory rather than definitive.

Attrition

The observed attrition at the child level (with respect to the primary outcome) was 21.6%: 25.3% in the intervention group and 17.8% in the control group (see Table 15). At setting level, one setting withdrew during the training in Cohort 2. At child level, between randomisation and data analysis stage, of 902 children randomised, 174 children were lost at endline assessment due to setting withdrawal (10), attendance absent (44), unable to engage/non-responsive (50), and left area or setting (64), and unknown reasons (6); 21 children were not included in analysis due to missing baseline data.

Table 15: Child-level attrition from the trial (primary outcome)

		Intervention	Control	Total
Number of children	Randomised	463	439	902
	Analysed	346	361	707
Child attrition (from randomisation to analysis)	Number	117	78	195
	Percentage	25.3%	17.8%	21.6%

Child and school characteristics

Information for baseline characteristics in Table 16 indicates that the control and intervention groups were broadly comparable at both the setting and child levels. Across the 101 participating settings, the distribution by cohort, nursery type, and Ofsted rating was similar, with only small differences such as a slightly higher proportion of Outstanding ratings in the intervention group and marginally greater deprivation (IDACI 1–3) among control settings. At the child level, the two groups were well balanced in terms of gender, EAL status, parental qualifications, and nursery attendance patterns. Although the control group had a slightly higher proportion of children with complete BPVS data and a greater share living in more disadvantaged areas, these differences were modest. Baseline attainment measures, including BPVS and ELIM raw scores and assessment ages, were closely aligned across groups, suggesting that children entered the study with similar starting points. Overall, the data show no substantive imbalances likely to affect the comparability of trial outcomes.

Table 16: Baseline characteristics of groups as randomised

Setting-level characteristics				
	Control (*n* = 50)		Intervention (*n* = 51)	
	n/N	%	*n/N*	%
Overall	50	49.5	51	50.5
Cohort				
Cohort 1 2023/2024	21/50	42.0	22/51	43.1
Cohort 2 2024/2025	29/50	58.0	29/51	56.9
Type of Nursery				
Maintained infant or primary school	22/50	44.0	22/51	43.1
Private for-profit nursery	24/50	48.0	26/51	51.0
Voluntary	4/50	8.0	3/51	5.9
Ofsted Rating				
Good	31/50	62.0	27/51	52.9
Not rated	2/50	4.0	1/51	2.0
Outstanding	17/50	34.0	23/51	45.1
Setting IDACI (1-3)	28/50	56.0	25/51	49.0

Child-level baseline characteristics					
	Control (*n* = 443)		Intervention (*n* = 453)		
Child level categorical	*n/N*	%	*n/N*	%	
Gender					
Female	228/443	51.47	226/453	49.89	
Male	212/443	47.86	224/453	49.45	
Unknown*	3/443	0.68	3/453	0.66	
EAL status					
No	320/443	72.23	323/453	71.30	
Yes	117/443	26.41	125/453	27.59	
Unknown*	6/443	1.35	5/453	1.10	
Mother qualification					
Primary school	2/443	0.45	9/453	1.99	
Secondary school up to 16	27/443	6.09	32/453	7.06	
Higher education (for example, BA Hons)	168/443	37.92	153/453	33.77	

Post-16 or further education (A Level, BTEC etc.)	70/443	15.8	77/453	17.0	
Post-graduate degree	124/443	27.99	128/453	28.26	
Prefer not to say/not applicable	32/443	7.23	33/453	7.28	
Unknown*	20/443	4.51	21/453	4.64	
Father qualification					
Primary school	4/443	0.90	6/453	1.32	
Secondary school up to 16	40/443	9.03	54/453	11.92	
Higher education (for example, BA Hons)	128/443	28.90	137/453	30.24	
Post-16 or further education (A Level, BTEC etc.)	77/443	17.38	77/453	17.00	
Post-graduate degree	117/443	26.41	97/453	21.41	
Prefer not to say/not applicable	52/443	11.74	56/453	12.36	
Unknown*	25/443	5.64	26/453	5.74	
Child Nursery hours					
Less than 15 hours per week	50/443	11.29	61/453	13.47	
15–20 hours per week	201/443	45.37	209/453	46.14	
21–30 hours per week	64/443	14.45	80/453	17.66	
More than 30 hours per week	107/443	24.15	77/453	17.0	
Unknown*	21/443	4.74	26/453	5.74	
Child IDACI (1 to 3)	196/443	44.2	158/453	34.9	
Child level continuous	*n/N*	mean(*sd*)	*n/N*	mean(*sd*)	Effect size (95% CI)
Child IDACI score	423/443	4.28(2.27)	428/453	4.86(2.46)	
Baseline BPVS Raw score	338/443	22.21(12.07)	316/453	22.52(11.28)	-0.03(-0.18, 0.12)
Baseline ELIM Raw score	432/443	33.06(14.08)	431/453	34.33(14.08)	-0.09(-0.22, 0.04)
Baseline BPVS Age in months	441/443	29.5(3.67)	450/453	30.06(3.56)	
Baseline ELIM Age in months	441/443	30.2(3.69)	450/453	30.66(3.61)	

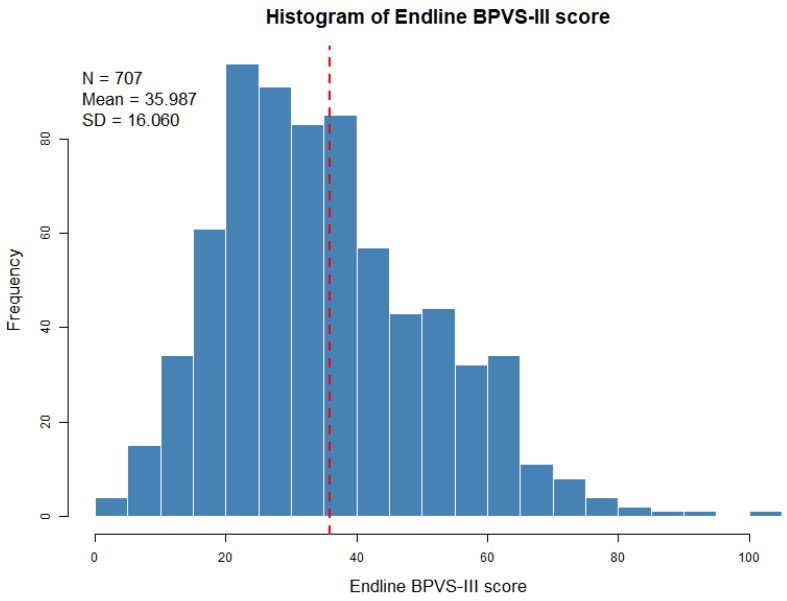
Note: *Unknown included all missing cases.

Outcomes and analysis

Primary analysis

Originally, the primary outcome would have been a combination of two language scores (BPVS-III and CDI-III) into a single score using the principal component analysis. However, since more than 20% of the first principal component of the combined BPVS-III and CDI-III data was missing (38.84% missing), we used only the BPVS-III score as the primary outcome. ELIM showed a moderate positive correlation with endline BPVS (r = 0.48, 95% CI [0.42, 0.53], p < .001).

Figure 3: Distribution of the endline BPVS-III score for whole sample



Overall across all children, the mean of the endline BPVS-III score was 35.99 with *SD* = 16.06. The primary outcome analysis shows no effect of the intervention on children's receptive vocabulary as measured by BPVS-III score at post-test (Hedges' *g* = -0.03; 95% CI: [-0.24, 0.18], *p* = 0.77; see Table 17), with the unadjusted intervention group mean of 36.32 (*SD* = 15.94) compared to the control group mean of 35.67 (*SD* = 16.29). The effect size estimate is negative but not statistically significant as evident from the *p*-value and 95% CI which included zero. The parameters used to estimate this effect are reported in Appendix C. The results of this analysis indicate that there was no difference between the receptive vocabulary of those children in EYCP settings compared to children in the control group across the full sample of children. This was not in line with the theory of change which expected the EYCP programme to positively impact on children's vocabulary skills.

Table 17: Primary outcome analysis results

	Unadjusted means				**Effect size**		
Outcome	Intervention		Control				
	N (missing)	Mean (95% CI)	*N* (missing)	Mean (95% CI)	Total *n* (intervention; control)	Hedges *g* (95% CI)	*p*-value
Endline BPVS-III	346 (85)	36.32 (34.64, 38.00)	361 (71)	35.67 (33.99, 37.35)	707 (346; 361)	-0.03 (-0.24, 0.18)	0.77

Secondary analysis

The secondary outcome analysis shows a small positive effect of the intervention on CDI-III total raw scores at post-test (Hedges' *g* = 0.03; 95% CI: -0.16, 0.21; see Table 18), with unadjusted mean scores for the intervention group of 66.04 and for the control group of 62.68; however, this result was not statistically significant (*p* = 0.58). The confidence interval includes zero, indicating that the true effect could range from slightly negative to slightly positive. Given the lack of statistical significance and the effect size estimate centred close to zero, results for the expressive vocabulary should be interpreted with caution, as it may reflect random variation rather than a true effect. The parameters used to estimate this effect are reported in Appendix C. Figure 4 shows the distribution of scores for this measure.

The secondary outcome WellComm shows a negative effect of the intervention at post-test (Hedges' *g* = -0.10; 95% CI: -0.30, 0.11; *p* = 0.47; see Table 19). The result should be interpreted with caution since it is not statistically significant. The unadjusted mean scores were very similar between the intervention and control groups. The intervention group had a mean score of 5.30 (95% CI: 5.05, 5.56) based on 350 children (103 missing), while the control group's mean score was 5.43 (95% CI: 5.18, 5.68) based on 361 children (82 missing). Figure 5 shows the distribution of scores for the WellComm measure. This indicates that the EYCP intervention did not impact on children's holistic language skills. These outcomes are not in line with the theory of change model that expected the EYCP intervention to positively impact on children's language skills.

Figure 4: Distribution of endline CDI-III score for whole sample

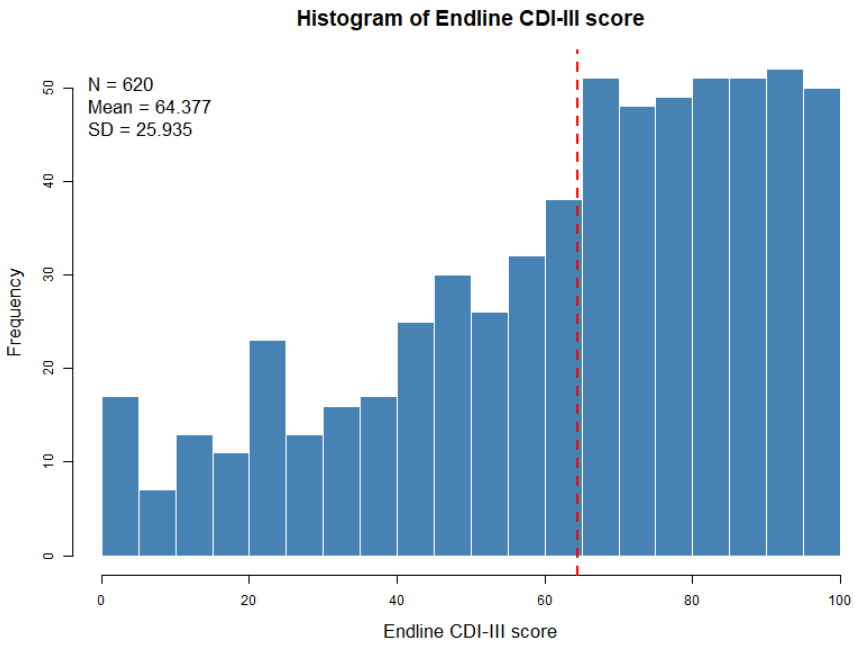


Table 18: Secondary outcome endline CDI-III score analysis results

	Unadjusted means				Effect size		
Outcome	Intervention		Control				
	N (missing)	Mean (95% CI)	*N* (missing)	Mean (95% CI)	Total *n* (intervention; control)	Hedges' *g* (95% CI)	*p*-value
Endline CDI-III score	314 (117)	66.04 (63.10, 68.97)	306 (126)	62.68 (59.83, 65.52)	620 (314; 306)	0.03 (-0.16, 0.21)	0.58

Figure 5: Distribution of WellComm score for whole sample

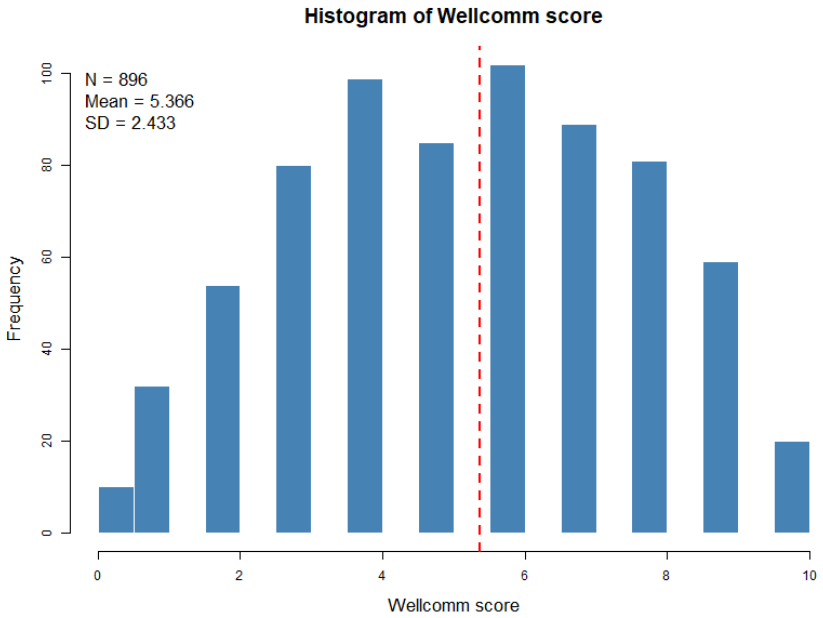


Table 19: Secondary outcome WellComm analysis results

	Unadjusted means				Effect size		
Outcome	Intervention		Control				
	N (missing)	Mean (95% CI)	*N* (missing)	Mean (95% CI)	Total *n* (intervention; control)	Hedges' *g* (95% CI)	*p*-value
WellComm	350 (103)	5.30 (5.05, 5.56)	361 (82)	5.43 (5.18, 5.68)	711 (350; 361)	-0.10 (-0.30, 0.11)	0.47

For the OLP practitioner outcome measure, across the three outcomes—Naming, Interpretation, and Perceiving —the unadjusted means show that practitioners in the intervention group scored higher than those in the control group across all three outcome scores. Figure 6, Figure 7, and Figure 8 show the distribution of scores on these measures.

For Naming, the control group had a mean score of 0.82 (95% CI: 0.19, 1.46) compared with 1.65 (95% CI: 0.76, 2.54) in the intervention group. The effect size (Hedges' *g* = 0.53; 95% CI: –0.06, 1.11) is not statistically significant, but suggests a moderate difference for practitioners in the EYCP intervention group.

For the Interpretation outcome, the intervention group mean was 3.65 (95% CI: 1.57, 5.72) versus 1.12 (95% CI: 0.42, 1.82) in the control group. The effect size (*g* = 0.66; 95% CI: 0.07, 1.25) is in the moderate range, and the confidence interval does not include zero, indicating a statistically significant difference. Given the number of comparisons we have made, it would be unwise to give any weight to this significant result.

For Perceiving, the intervention group mean was 61.9 (95% CI: 54.0, 69.7) compared with 52.4 (95% CI: 42.7, 62.1) in the control group. The effect size (*g* = 0.45; 95% CI: -0.13, 1.04) confidence interval includes zero, indicating no statistically significant difference between the practitioner outcome in the intervention and control groups.

Overall, practitioners (EYCP Champions) in the intervention group demonstrated consistently higher scores across all three outcomes, suggesting a positive effect of the EYCP on language-related pedagogical knowledge. However, the lack of

statistical significance for two of the three outcomes (Naming and Perceiving), alongside wide confidence intervals, likely reflects limited statistical power due to the small sample size. As a result, while the direction of effects is consistently positive, confidence in the precision of these estimates is limited, and the findings should be interpreted with caution. These findings are broadly consistent with the EYCP Theory of Change, which posits that participation in the programme would enhance practitioners' knowledge and application of language pedagogy.

Figures 6 to 8 show that there is some clustering of scores at the lower end of the scale (floor effects) in the Naming and Interpretation subtests which may indicate some deviation from normality distribution. However, formal tests of normality were not conducted, as these tests have limited reliability and low power in small samples, and may fail to detect meaningful deviations from normality. Linear models used for OLP score analysis are generally robust to moderate departures from normality, particularly when estimating differences between groups. However, the presence of floor effects and skewed distributions may reduce variability and contribute to wider confidence intervals, limiting the precision of estimates. These factors, together with the small sample size, should be considered when interpreting the findings.

Table 20: OLP score analysis results

	Unadjusted means				**Effect size**		
OLP Outcome	Intervention		Control				
	N (missing)	Mean (95% CI)	*N* (missing)	Mean (95% CI)	Total *n* (intervention; control)	Hedges' *g* (95% CI	*p*-value
Naming	34(17)	1.65 (0.76, 2.54)	17(33)	0.82 (0.19, 1.46)	51(34; 17)	0.53 (-0.06, 1.11)	0.06
Interpretation	34(17)	3.65 (1.57, 5.72)	17(33)	1.12 (0.42, 1.82)	51(34; 17)	0.66 (0.07, 1.25)	0.02
Perceiving strategies	34(17)	61.9 (54.0, 69.7)	17(33)	52.4 (42.7, 62.1)	51(34; 17)	0.45 (-0.13, 1.04)	0.12

Figure 6: Distribution of the OLP Naming scores for the whole sample

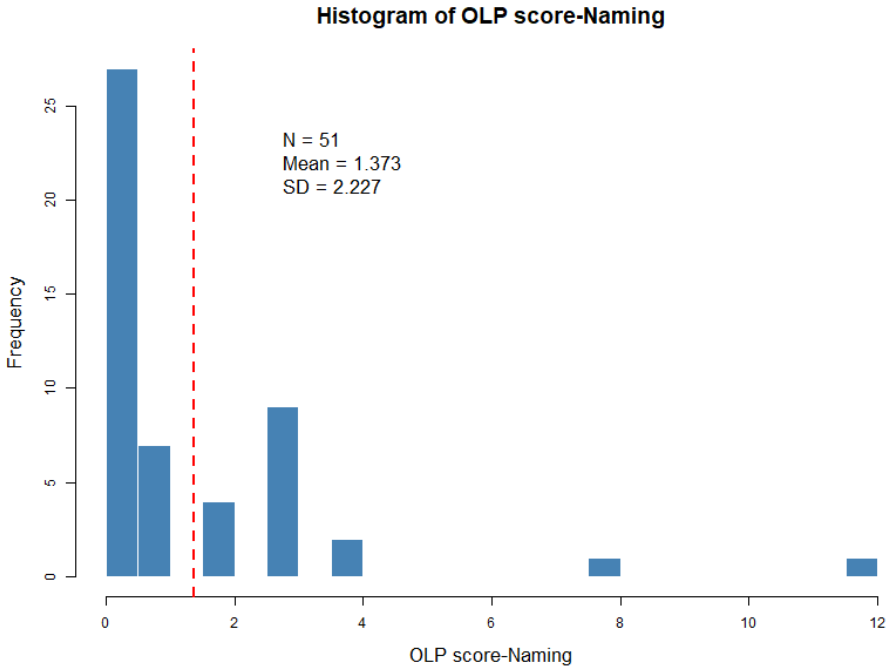


Figure 7: Distribution of the OLP Interpretation scores for the whole sample

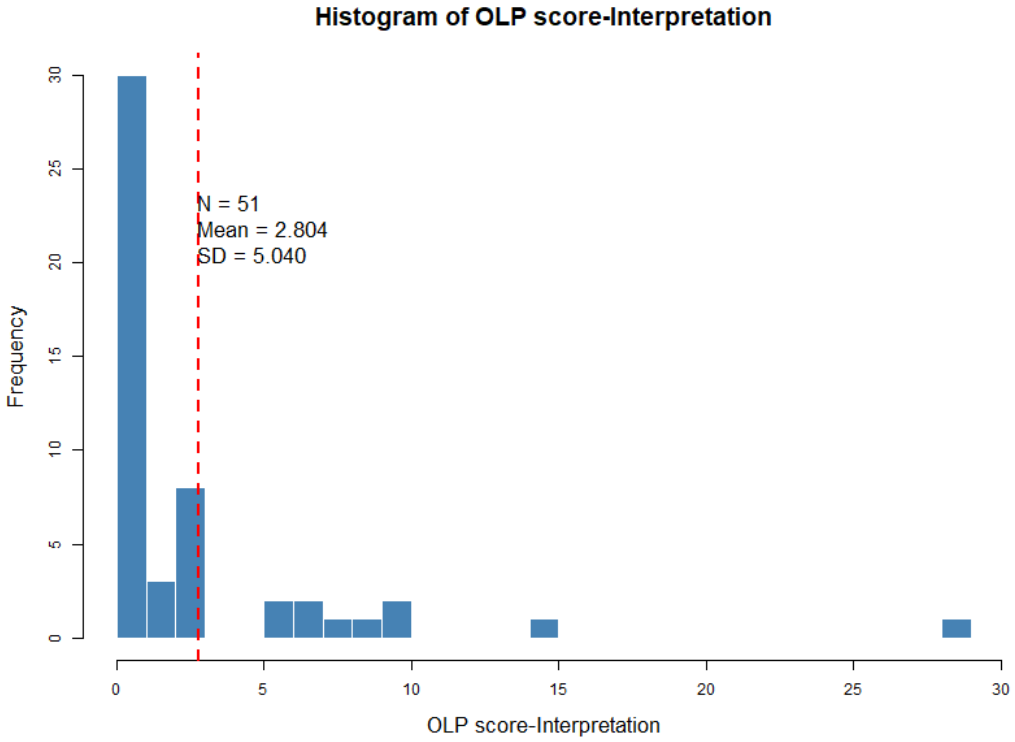
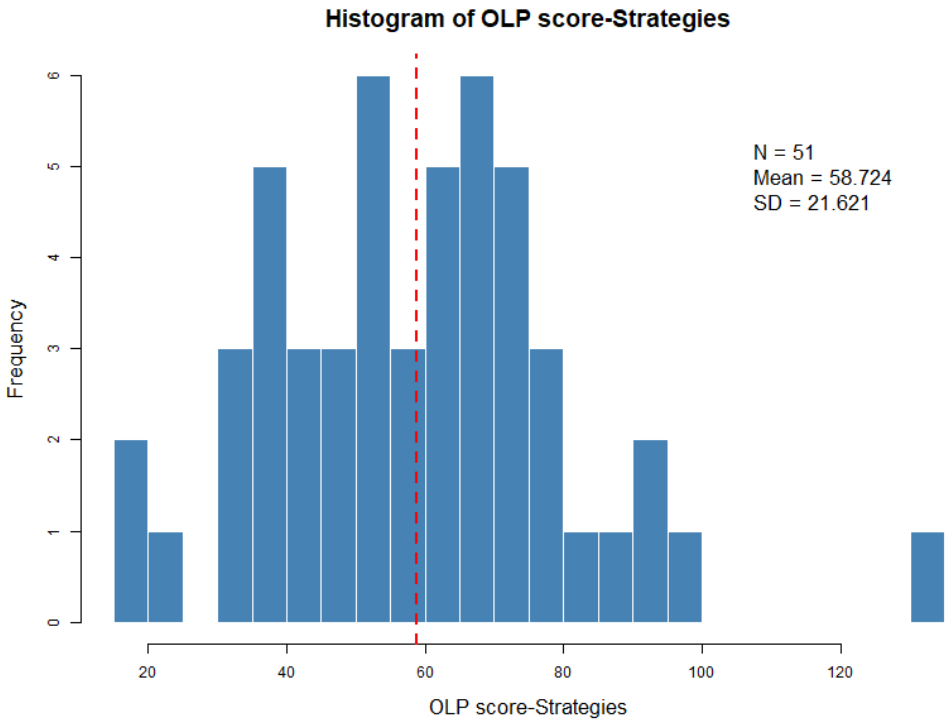


Figure 8: Distribution of the OLP Perceiving scores for the whole sample



Analysis in the presence of non-compliance

Three sets of CACE analyses were conducted to assess 1) the effect of overall compliance considering setting- and child-level compliance, 2) minimal compliance for both setting and child, and 3) optimal compliance for both setting and child. Minimal compliance was observed for 262 children across 52 nurseries, while optimal compliance was observed for 49 children across 28 nurseries. When looking at:

1) overall compliance, each one unit increase in compliance (representing the contrast between no and full combined compliance) was associated with a +6.76 point increase in the intervention effect on endline BPVS-III (95% CI: 1.23 to 12.28). This corresponds to a compliance-adjusted effect size (Hedges' *g*) of 0.47 (95% CI: 0.26 to 0.69) with rising compliance. It is also important to highlight that compliance here is a rescaled variable from 0 to 1; therefore, this effect size estimate reflects the full contrast between the lowest and highest observed levels of engagement and therefore represents a cumulative dose-response effect across the compliance spectrum. This result, however, indicates a positive relationship between greater engagement in the elements of the EYCP programme and positive child language outcomes. The overall compliance measure combines setting-level fidelity with child-level dosage/exposure into a single rescaled metric. As such, the estimated effect reflects a composite of increased engagement at the setting level, increased exposure at the child level, or both. Consequently, the positive association observed cannot be attributed to either component independently, and the relative contribution of setting fidelity versus child exposure remains unclear. This makes the interpretation of the overall compliance effect more complex. The estimate should be understood as capturing a joint (and potentially heterogeneous) compliance construct, rather than a single, well-defined intervention contrast. Some additional exploratory analysis for the intervention effect and compliance by child-level and setting-level compliance separately shows that there was a positive and statistically significant interaction between the intervention and child-level compliance. In contrast, there was no evidence of any significant interaction between the intervention and setting-level compliance, suggesting that setting-level compliance did not modify the intervention effect.

Or 2) minimal compliance, the additional intervention effect among minimal compliers (both setting and child level) vs nonminimal was -5.54 (SE 7.25, 95% CI: -19.75 to 8.67) i.e., not statistically significant, and the compliance-adjusted effect size was -0.38 (95% CI: -0.62 to -0.15). Where settings were measured as minimally compliant based on a minimally

acceptable level of the EYCP programme and children measured as compliant based on the minimum level of exposure specified as part of the inclusion criteria, children in these settings did not score more highly on the language outcome measure indicating that minimal compliance is not sufficient for change to happen to child-level outcomes.

For 3) optimal compliance, the additional intervention effect among optimal compliers (both setting and child level) vs others is +1.47 (SE: 2.49, 95% CI: -3.41 to 5.88) and the effect size for the group who complied optimally was 0.10 (95% CI: -0.13 to 0.34). Where settings were measured as optimally compliant based on an optimal acceptable level of the EYCP programme and children measured as optimally compliant based on the maximum level of exposure specified as part of the inclusion criteria, there were positive impacts of the programme on children's language outcomes. These results indicate that increased engagement with all EYCP programme activities led to greater effects of the programme.

Missing data analysis

From Table 17, it is already evident that 85 cases were missing from the intervention group while 71 cases were missing from the control group, which was not significantly different. However, an adjusted mixed-effect logistic regression with presence or absence of post-test BPVS-III score as the outcome variable was conducted to examine if the missing post-test scores were related to any observed characteristics including baseline ELIM scores, intervention, cohort, and stratification factors (type of setting, median number of children, local authority). ELIM score at pre-test was the only covariate observed to be a statistically significant predictor of the missing post-test scores, assessed at the 5% level indicating the missingness could be at random (MAR). Logistic regression helps assess whether missingness depends on observed data. However, residual dependence of missingness on unobserved outcomes cannot be completely ruled out and, if present, may bias results due to non-random missing data. Pre-test ELIM score was then included in the multiple imputation model for the missing endline BPVS-III. To investigate the impact of missing data, the primary analysis was repeated using predicted values obtained via multiple imputation by chained equations (MICE).

The adjusted mean difference in BPVS-III score between the intervention and control groups following multiple imputation was -0.56 (95% CI: -2.84, 1.72), which is not statistically significant at the 5% significance level. The corresponding estimated effect size for the intervention effect is -0.04 standard deviations (95% CI: -0.23 to 0.16), which is similar to the results of the primary analysis increasing confidence in the primary analysis despite the level of missing data. The total variance used to calculate the effect size was 255.63—the sum of 222.73 (random variation between children) and 32.9 (heterogeneity between nurseries, between-cluster variance).

Subgroup analyses

The main analysis was also conducted on several subsamples of children: a) those who had English as an additional language; b) the top (20%) and bottom (20%) performers as measured from the quantile of the baseline scores (ELIM), and c) the children whose residence is in an IDACI 1–3 postcode. Subgroup analyses of the BPVS primary outcome showed no evidence that the intervention effect differed meaningfully across the examined child characteristics (see Table 21). For children with English as an additional language (EAL), the effect size was -0.05 (95% CI: -0.40, 0.30), indicating no significant difference between the intervention and control EAL children. Similarly, among children in the bottom 20% of baseline performance, the effect size was -0.28 (95% CI: -0.70, 0.13), and among those in the top 20%, the effect size was -0.17 (95% CI: -0.54, 0.20). For children living in areas ranked IDACI 1–3, the effect size was -0.11 (95% CI: -0.43, 0.22). Although these point estimates are negative and suggest a tendency towards lower progress in the intervention group across subgroups, the confidence intervals crossed zero, suggesting that the intervention could not detect differential effects for any of these specific subgroups. This also indicates that the EYCP intervention did not demonstrate statistically significant impact on children's language skills (particularly receptive vocabulary) for these subgroups. Consistent with the EEF statistical guidance, these subgroup analyses should be interpreted as exploratory, and findings should be considered with appropriate caution given the increased likelihood of Type I error arising from multiple comparisons. It is also noted that the subgroup analyses did not yield statistically significant effects, which further reduces the risk of over-interpreting spurious findings.

For the secondary outcome (CDI-III total score which was initially proposed as part of the combined primary outcome), effect sizes across all subgroups were small and not statistically significant. For children with EAL, the effect size was 0.04 (g = 0.04, CI: -0.28, 0.35) but not statistically significant, indicating no difference between the intervention and control groups. Among children in the top 20% of baseline performance, an effect size of 0.18 was observed (g = 0.18, CI: -0.19, 0.56), suggesting a slight tendency for higher-performing children to benefit; however, the confidence interval includes zero, indicating that this difference is not statistically reliable and should be interpreted with caution (see Table 22). For children living in areas ranked IDACI 1–3, the effect size was -0.01 (95% CI: -0.28, 0.25).

Table 21: BPVS primary outcome analysis by subgroup

	Unadjusted means				**Effect size**		
BPVS Outcome	Intervention		Control				
	N (missing)	Mean (95% CI)	*N* (missing)	Mean (95% CI)	Total *n* (intervention; control)	Hedges' *g* (95% CI	*p*-value
EAL	90 (27)	29.28 (26.65, 31.91)	95 (18)	28.89 (25.98, 31.81)	185 (90; 95)	-0.05 (-0.40, -0.30)	0.93
Bottom 20%	49 (33)	23.04 (19.67, 26.41)	71 (23)	23.99 (21.61, 26.37)	120 (49; 71)	-0.28 (-0.70, 0.13)	0.17
Top 20%	68 (12)	43.62 (39.74, 47.49)	67 (4)	47.61 (43.44, 51.78)	135 (68; 67)	-0.17 (-0.54, 0.20)	0.32
IDACI 1 to 3	108 (39)	32.84 (30.00, 35.68)	156 (35)	34.14 (31.51, 36.77)	264 (108; 156)	-0.11 (-0.43, 0.22)	0.56

Table 22: CDI-III secondary outcome analysis results by subgroup

	Unadjusted means				**Effect size**		
CDI-III Outcome	Intervention		Control				
	N (missing)	Mean (95% CI)	*N* (missing)	Mean (95% CI)	Total *n* (intervention; control)	Hedges' *g* (95% CI	*p*-value
EAL	76 (41)	61.74 (55.49, 67.98)	80 (33)	58.80 (52.93, 64.67)	156 (76; 80)	0.04 (-0.28, 0.35)	0.46
Bottom 20%	53 (29)	31.62 (24.54, 38.71)	65 (29)	35.25 (29.40, 41.09)	118 (53; 65)	-0.13 (-0.52, 0.25)	0.30
Top 20%	60 (20)	86.05 (82.61, 89.49)	52 (19)	80.79 (75.19, 86.38)	112 (60; 52)	0.18 (-0.19, 0.56)	0.39
IDACI 1 to 3	99 (48)	63.56 (57.67, 69.44)	124 (67)	60.94 (56.08, 65.81)	223 (99; 124)	-0.01 (-0.28, 0.25)	0.76

Additional analyses and robustness checks

Difference between Cohort 1 and Cohort 2

The difference between Cohort 1 and 2 was assessed by fitting an interaction model with intervention and cohort and checking if there is a significant interaction effect based on *p*-values. Table 23 shows the results from an additional cohort-specific analysis of the primary outcome conducted for each cohort separately as a robustness check. Overall, the intervention does not appear to have a meaningful effect in either cohort. However, the effect size remains negative and statistically non-significant for both the cohorts, which is consistent to what we found for the overall primary outcome analysis.

Table 23: Primary outcome results by cohort

	Intervention		Control		Effect size		
BPVS Outcome	*N* (missing)	Mean (95% CI)	*N* (missing)	Mean (95% CI)	Total *n* (Intervention; control)	Hedges' *g* (95% CI)	*p*-value
Cohort 1	139 (39)	33.13 (30.57, 35.68)	150 (34)	32.30 (29.87, 34.73)	289 (139; 150)	-0.08 (-0.36, 0.20)	0.72
Cohort 2	207 (46)	38.46 (36.27, 40.65)	211 (37)	38.06 (35.80, 40.32)	418 (207; 211)	-0.04 (-0.33, 0.24)	0.36

Exclusion based on nursery attendance

A sensitivity analysis of the primary outcome results was conducted excluding children who received the intervention despite not fulfilling the eligibility criteria of attending nursery for more than 15 hours per week from the full sample. The effect size was not different after excluding the cases with less than 15 hours of attendance (effect size -0.05 (-0.26, -0.17)) compared to the primary outcome analysis.

Differences by baseline measures

To assess the effect of using different baseline outcomes on primary outcomes results, additional analysis was conducted replacing the ELIM baseline score with the BPVS-III baseline score and redoing the primary outcome analysis. This analysis was conducted using the BPVS-III baseline score only for the sample with both ELIM and BPVS-III scores, to aid interpretation of any differences in the results. It was found that the effect size for the primary outcomes was negative and like the main primary outcome analysis (effect size -0.05; CI: -0.29, - 0.20). However, the low data collection rates for BPVS-III limit the reliability of this comparison.

Additional analysis for secondary outcome

Additional secondary outcome mixed-effects logistic regression analysis was conducted using the WellComm assessment RAG rating as a binary outcome variable to look at whether there were fewer children needing additional support as a result of the intervention (i.e., Green = no intervention needed, Amber/Red= additional support is needed and/or consider referral to a Speech and Language Therapist). The intervention group has about 7% lower odds of a child needing additional support than the control group. The effect size estimate is -0.4 (95% CI: -0.27, 0.18).

Additional subgroup analysis

Due to no measure of FEE status eligibility in Cohort 2, additional analysis was conducted for the FEE at age two subgroup for Cohort 1 and the IDACI 1–3 subgroup for Cohort 1. This analysis allowed examining the variation in the effect size by considering two different sets of disadvantage indicators (FEE vs IDACI). The effect size for the primary outcome for the FEE

subgroup was -0.04 (95% CI: -0.36, 0.29). Similarly, the effect size for the IDACI 1–3 subgroup was -0.03 (95% CI: -0.40, 0.34). These findings indicate that the results for the FEE and IDACI 1–3 subgroups were similar and comparable for cohort 1.

An analysis for the primary outcome by age was conducted to examine the sensitivity of the results towards younger and older children who may have responded differently. The primary outcome effect size was not similar for both the age groups (i.e., age less than equal to 30 months had the effect size of -0.09 (95% CI: -0.32, 0.13); and the age greater than 30 months had also the effect size of 0.04 (95% CI: -0.27, 0.35)).

Additional analysis for the primary outcome by staff level of qualification showed no significant effect of intervention. For the subgroup of children taught by staff with qualification level 4 or below, the intervention effect size was -0.08 (95% CI: -0.31, 0.15), while for children taught by staff with qualification level 5 or above, the effect size was -0.17 (95% CI: -0.56, 0.22). Although the estimated effect size for the higher-qualified staff group was more than twice as negative as that for the lower-qualified staff group, the small sample sizes reduce confidence in these findings. Consequently, neither effect is statistically significant, and there is insufficient evidence to conclude that the intervention's impact differed between the two groups.

These findings indicate that the intervention effect did not vary meaningfully by staff qualification level.

ICC

ICC was estimated using a two-level multilevel model of endline BPVS-III scores without any covariates (the empty model) and the model containing all covariates proposed for the BPVS-III primary outcome analysis. The unconditional ICC for the primary outcome was 0.15. When controlling for all covariates and the clustering of children in nurseries using the primary analysis model, the ICC for the primary outcome was 0.12.

Implementation and process evaluation results

Key findings

Findings from the implementation and process evaluation showed strong engagement and high satisfaction with EYCP among nursery staff. Practitioners reported increased confidence, improved pedagogical knowledge, and observable shifts in practice which included greater use of the ShREC approach and interactive book reading in daily interactions. Barriers to being able to fully engage with EYCP included being able to release staff, high workload, geographical constraints in getting to training and mentoring sessions, and challenges working with two-year-olds. Facilitators to delivering the programme well included strong mentor support, accessible training resources, leadership buy-in, and opportunities for video-based reflection. It was found that both intervention and control settings already engaged in a range of language-related practices and CPD activities prior to and during the intervention period. The intervention was implemented within contexts where existing language-support practices were already embedded to varying degrees. This may partly explain improvements observed within some control settings and highlights the challenge of distinguishing intervention effects from ongoing usual practice. The intervention was considered feasible to implement across early years settings with manageable challenges while working with the targeted 2-year-olds group in outcome assessments, although implementation burden varied across contexts. Despite these limitations, the evaluation provided evidence that the intervention could be delivered at scale with appropriate support structures, monitoring systems, and leadership engagement.

Compliance

Summary of findings

- Setting training attendance data for two key members (EYCP Champion and EY lead/manager) was collected for 51 out of 52 settings in the intervention group which includes 29 PVI settings and 22 SN settings.
- Compliance at setting level was considered high for both managers' (mean 2.47 out of 3 sessions) and EYCP Champions' attendance (mean 19.08 out of 22 sessions) across the EYCP intervention programme.
- PVIs and SNs settings' compliance to EYCP training were similar with high levels of attendance, although SNs had slightly higher attendance (manager/Champions: 2.50/19.27) than PVI settings' attendance scores (manager/Champions: 2.45/18.93).
- There were found no significant setting-level attendance difference by study cohorts or setting's IDACI status across hubs.

In this trial, compliance is defined as 'the extent to which the critical ingredients of the EYCP intervention are delivered to and/or received by the target participants (settings/practitioners/children)'. In the impact evaluation, the setting compliance indicator was represented as a continuous variable in the form of the core components of the EYCP intervention (0–10), as listed in Table 4.

In the IPE, we further explore setting compliance in relation to two key members' (EYCP Champion and setting manager) attendance records across EYCP programme including full-day and half-day in-person training events, online PD modules, hub meetings, and mentor visits. See below for the breakdown of each training component.

- Face-to-face training events: two (one full day at the start of training and one half day at the midpoint)
- Six online PD training modules
- Nine Hub meetings: (alternating in person (five) and online (four))
- Five mentoring visits to setting: (two to conduct ITERS evaluation at baseline and endline, three support visits throughout the training).

This section partially addresses the IPE research question RQ1: 1.1, 1.4.

Data used to answer this question has been synthesised from the administrative records provided by the DT which kept records of EYCP training attendance and reasons for absence for both managers and Champions for each setting.

RQ1: Is the EYCP PD training and mentoring support delivered as intended to PVI, maintained, and SN settings?

1.1. **Are nominated staff (manager/EYFS lead, EYCP Champion) attending EYCP training and the hub meetings (EYCP Champion) as specified in the programme plan?**

1.5. To what extent does the EYCP programme and its delivery to settings in Cohort 1 differ from the delivery in Cohort 2?

EY leads/managers' attendance to PD training and hub meetings (Administrative records)

As specified in the programme, managers were expected to attend the two in-person training sessions and one online module. Attendance is reported for each session (each scored 0: not attended or 1: attended) and a total attendance score (maximum score of 3). Managers were not expected to attend mentors' support visits in setting. Reasons for absence were noted down by the DT.

As shown in Table 24, managers' attendance was high across all training activities, with mean scores of .80 for launch-day in-person training, .78 for the half-day in-person training, and .88 for the online training module.

Managers' mean total attendance score from a possible 3 was 2.47 demonstrating good adherence to the programme overall.

EYCP Champions' attendance to PD training, hub meetings, and mentors' support visits (Administrative records)

Champions were expected to attend all sessions and host all support visits. Attendance is reported for the two in-person training sessions (each scored 0/1), a summed score for attendance at online modules (maximum 6), a summed score for attendance at hub meetings (maximum 9), a summed score for hosting mentor visits (maximum 5), and a total attendance score (maximum 22). Reasons for absence were noted down by the DT.

As shown in Table 24, Champions' attendance in person was slightly higher than managers' at 92% and 90% for the launch day and half-day training events respectively. Champions had a high level of attendance in all other strands of the training, attending an average of 5.06 online modules out of a possible six, 7.37 hub meetings out of a possible nine, and hosting 4.82 mentor visits from a maximum of five (only the two settings that did not complete the training failed to host all mentor meetings). Overall Champions received a mean total attendance score of 19.08 (*SD* = 3.8) out of 22.

Table 24: Attendance at EYCP training and support

Manager and Champion attendance: combined cohorts					
	N	Min	Max	Mean	Std. Deviation
Manager Launch day attendance	51	0	1	.80	.401
Manager Half day attendance	51	0	1	.78	.415
Manager module 1 attendance	51	0	1	.88	.325
Manager total attendance score	51	1	3	2.47	.674
Champion Launch day attendance	51	0	1	.92	.272
Champion Half day attendance	51	0	1	.90	.300
Champion online module attendance	51	1	6	5.06	1.207

Champion hub meeting attendance	51	0	9	7.37	2.029
Setting visits completed	51	0	5	4.82	.888
Champion total attendance score	51	1	22	19.08	3.810

Difference in training attendance across cohorts

As shown in Table 25, managers' total training attendance score in Cohort 1 was slightly higher than Cohort 2 (2.55 compared with 2.41), but they were not significantly different, $t(49) = .687$, $p = .495$. The greatest difference was with the launch day in-person training in which 86% (mean=.86) of managers attended during Cohort 1 and 76% (mean .76) in Cohort 2.

In contrast, Champions had slightly higher overall attendance scores in Cohort 2 compared to Cohort 1 (19.28 compared to 18.82), although the difference was small and not statistically significant, $t(49) = -.427$, $p = .675$. The greatest difference between the two cohorts was attendance at the half-day in-person meeting (.93 vs .86) and attendance at online training modules (5.24 vs 4.82).

Table 25: Difference in attendance at training by cohort

Manager and Champion attendance by cohort						
Cohort		***N***	**Minimum**	**Maximum**	**Mean**	**Std. Deviation**
Cohort 1	**Manager Launch day attendance**	22	0	1	.86	.351
	Manager Half day attendance	22	0	1	.77	.429
	Manager module 1 attendance	22	0	1	.91	.294
	Manager total attendance score	22	1	3	2.55	.596
	Champion Launch day attendance	22	0	1	.95	.213
	Champion Half day attendance	22	0	1	.86	.351
	Champion online module attendance	22	1	6	4.82	1.296
	Champion hub meeting attendance	22	1	9	7.36	1.891
	Setting visits completed	22	1	5	4.82	.853
	Champion total attendance score	22	4	22	18.82	3.647
	Valid *N* (listwise)	22				
Cohort 2	**Manager Launch day attendance**	29	0	1	.76	.435
	Manager Half day attendance	29	0	1	.79	.412
	Manager module 1 attendance	29	0	1	.86	.351
	Manager total attendance score	29	1	3	2.41	.733
	Champion Launch day attendance	29	0	1	.90	.310
	Champion Half day attendance	29	0	1	.93	.258
	Champion online module attendance	29	1	6	5.24	1.123

	Champion hub meeting attendance	29	0	9	7.38	2.162
	Setting visits completed	29	0	5	4.83	.928
	Champion total attendance score	29	1	22	19.28	3.981

Difference in attendance by setting type and setting IDACI status

As shown in Table 26, manager and EYCP Champion attendance scores were broadly similar regardless of type of setting. Both managers and Champions in maintained settings had slightly higher mean total scores compared to PVI settings but not significantly so for either managers, $t(49) = .269$, $p = .789$, nor Champions, $t(49) = .314755$, $p = .755$, and both setting types had high levels of attendance.

Settings were grouped into three categories based on the IDACI decile score indicated by their postcode (1–3, 4–6, 7+) to investigate how socio-economic status of the area impacted attendance. Managers' overall attendance increased slightly in line with higher IDACI scores, but the difference was small, and a One-way ANOVA revealed no significant difference in attendance depending on setting IDACI rating, $F(2, 48) = 1.852$, $p = .168$. For Champions, lowest attendance scores were found in the IDACI 4–6 group, and highest in the 7+ group, but again no significant differences were detected following a One-oay ANOVA, $F(2, 48) = .085$, $p = .919$.

Table 26: Difference in attendance by setting type

Manager and Champion attendance by setting type						
Setting Type		***N***	**Minimum**	**Maximum**	**Mean**	**Std. Deviation**
Private or Voluntary	**Manager total attendance score**	29	1	3	2.45	.736
	Champion total attendance score	29	1	22	18.93	3.909
Maintained	**Manager total attendance score**	22	1	3	2.50	.598
	Champion total attendance score	22	4	22	19.27	3.757

Fidelity

Summary of findings

- The EYCP PD training and mentoring support were perceived as highly appropriate, acceptable, and beneficial.
- Hub meetings and mentoring were delivered largely in accordance with the manualised structure, with minor adaptations to accommodate contextual needs.
- The main barriers to fidelity related to structural constraints (time, staffing, geography) and technical issues, rather than deficiencies in programme design. Strong mentor support, structured materials, video exemplars, and a blended learning model were key facilitators supporting high fidelity of training delivery.
- Across cohorts, fidelity of implementation was high: Champions clearly understood and adhered to their roles, and cascading occurred in most settings (87.8%).
- Variation in implementation fidelity was explained by leadership engagement; staffing capacity; setting structure and environment; Champions' confidence; staff turnover; setting-level socio-economic status (SES); and other contextual factors.

- Cohort 2 benefitted from refinements informed by Cohort 1, particularly increased clarity, earlier baseline assessment, and stronger focus on interactive reading. Despite increased mentor workload, continuity of the mentor team supported consistency and quality across cohorts.

Fidelity of training delivery

This section addresses the IPE research question RQ1: 1.2, 1.3, 1.4, 1.5.

Data used to answer this question has been synthesised from the practitioner endline survey part III: EYCP evaluation (*n* = 33), setting interview (*n* = 23), mentors' interview (*n* = 20), developers' interview (*n* = 4), and hub meeting observation (*n* = 28).

RQ1: 1.2. How appropriate are the PD training sessions (for example, content, coverage, dosage, and duration) and the level of support for nominated staff? Are PD training sessions and mentoring support (including hub meetings and setting visits) delivered as manualised?

Acceptability and Appropriateness of training

A total of 33 EYCP Champions (14 from Cohort 1 and 19 from Cohort 2) completed the endline evaluation survey (response rate 64.7%). The evaluation section of the survey consisted of 20 items scored on a 0–4 scale, divided into three categories: satisfaction with professional development (PD) training (online modules and in-person training); perceived importance of each PD training session (each individual online module and the in-person training); acceptability of training delivery (focusing on mentor support).

PD training

As shown in Table 27, EYCP Champions reported high levels of satisfaction with the EYCP content, delivery, dosage, benefit, and accessibility of the PD training, with mean scores ranging from 3.58 to 3.85. Of those 33 EYCP Champions, 31 Champions agreed that the EYCP training was delivered at the right level; 32 Champions agreed that the PD training sessions were accessible and delivered with appropriate content, frequency, and dosage; the training provided sufficient information and support to Champions to implement the EYCP in their practice, and it benefits their own professional development.

Table 27: Mean satisfaction scores on EYCP PD training

Key Elements of EYCP Training	*N*	Minimum	Maximum	Mean	Std. Deviation
Training delivered at the right level	33	2	4	3.61	.609
Training was accessible	33	2	4	3.70	.529
Training benefits professional development	33	2	4	3.85	.442
Training materials were accessible	33	2	4	3.76	.502
Training provided enough information and support to get started with EYCP	33	2	4	3.70	.529
The PD training was appropriate in frequency and length	33	2	4	3.58	.561

Perceived importance of PD training

Table 28 shows mean scores for the total sample regarding the perceived importance of the online and in-person PD training sessions. Each item is rated 0–4 with 0 being not important at all and 4 being very important. Results indicated that EYCP Champions perceived all of the PD training sessions to be important to them, with mean scores ranging from 3.70–3.91, with the online module on inclusive practice receiving the highest score (mean score: 3.91 out of 4), and the online module on multilingualism receiving the lowest with 3.70 out of 4. Most respondents felt all modules and the in-person training were important to some extent.

Table 28: Perceived importance of EYCP PD training

EYCP PD training	*N*	Minimum	Maximum	Mean	Std. Deviation
Importance of 1.5 days face-to-face session	33	1	4	3.73	.674
Importance of module 1: Approaches to supporting colleagues	33	2	4	3.79	.485
Importance of module 2: Inclusive practice	33	3	4	3.91	.292
Importance of module 3: Prioritising back and forth interactions	33	3	4	3.88	.331
Importance of module 4: Multilingualism	33	2	4	3.70	.585
Importance of module 5: Working with families	33	2	4	3.82	.465
Importance of module 6: Embedding and sustaining the ShREC Approach	33	2	4	3.82	.465

Hub meetings and mentors' support

Table 29 shows mean scores for the total sample regarding the accessibility, acceptability, and benefit of hub meetings, mentors' support visits, and the use of video reflections. All items are scored on a 0–4 scale with 4 being 'completely agree' (indicating the most positive response) and 0 being 'completely disagree' (the least positive). Mean scores for all items ranged from 3.15 to 3.94 indicating high levels of acceptability and perceived benefit of the training. The highest-rated aspects were the mentors' in-school support visits (M = 3.94, SD = 0.24), the perceived benefit of mentor support more broadly (M = 3.88, SD = 0.33), and the video reflection using the ShREC approach (M = 3.85, SD = 0.36). These findings suggest that personalised, practice-based support and reflective components were particularly valued by practitioners. The lowest mean score was regarding accessibility of the hub meetings (M = 3.15, SD = 0.97) suggesting some respondents may have struggled to attend or engage with these meetings. Overall, practitioners reported high levels of satisfaction with EYCP hub meetings and mentor support visits.

Table 29: Mean satisfaction scores on EYCP hub meeting and mentor support visits

	N	Minimum	Maximum	Mean	Std. Deviation
The time allocated for hub meetings was about right for me	33	2	4	3.48	.712
The frequency of the hub meetings was about right for me	33	1	4	3.58	.751
The level of mentor support was beneficial for me	33	3	4	3.88	.331
The accessibility of the hub meetings was easy	33	1	4	3.15	.972
The delivery mode of the hub meeting was well balanced	33	1	4	3.58	.792
The mentors' in-school support visits were beneficial for me	33	3	4	3.94	.242
The video reflection of my teaching using ShREC approach was beneficial for me	33	3	4	3.85	.364

The findings from the endline survey were largely consistent with those obtained from interviews and focus groups conducted with setting Champions and managers (n = 23), mentors (n = 20), and developers (n = 4).

Across all settings, participants reported a consistently positive experience of the EYCP training. This was reflected in qualitative feedback, with one participant noting: *'I definitely feel privileged to attend, to take part in it. It was incredible. I loved every minute of it, every minute'* (MS51A). An EYCP Champion from Cohort 1 commented that the training was:

> *'So accessible, we have everything on the Padlet, the videos we have on the WhatsApp group. And all the time we have a reminder about the meetings, and they send to us on the email as well.'* (CN11)

Champions and managers identified several key elements of the EYCP training as particularly valuable. These included the in-setting support and visits provided by mentors (*n* = 14), and the use of video alongside RNAR forms (Rehearse, Notice, Analyse, and Refine), which were perceived to facilitate reflective practice and enhance understanding of the impact of ShREC strategies on children's interactions (*n* = 10). Participants also highlighted the benefits of the programme's multi-strand design—comprising professional development (PD), hub meetings, and support visits—and its hybrid delivery model (combining online and in-person formats) (*n* = 7).

In addition, respondents considered the frequency, duration, and overall dosage of the training to be appropriate (*n* = 8). Opportunities for Champions to engage with peers, including meeting other practitioners and visiting different settings, were also highly valued (*n* = 8). The training content was regarded as relevant and well tailored to the early years context (*n* = 7), while specific resources, such as Padlet, were noted as useful (*n* = 5). A smaller number of participants highlighted the repetitive and consistent nature of the training as beneficial in reinforcing learning (*n* = 4).

Mentors reported similar perspectives during interviews, with themes closely aligning with those identified by Champions and managers, particularly in relation to the accessibility and acceptability of the EYCP training. Mentors emphasised settings' enjoyment of and positive engagement with the programme (*n* = 13), as well as the effective use of video as a training tool (*n* = 10). They also noted the relevance of the training to early years practice (*n* = 6), the value placed on peer interaction and cross-setting visits (*n* = 6), and the advantages of the multi-strand, hybrid approach (*n* = 4).

Furthermore, mentors highlighted additional strengths of the programme, including its responsiveness to feedback, with ongoing adaptations to training delivery (*n* = 12), and the clarity of its approach and messaging (*n* = 6). They also observed that the EYCP aligned well with practitioners' professional values (*n* = 4) and contributed to enhanced professional pride and self-worth (*n* = 4).

Approaches and adherence to training structures (hub meeting observation; DT interview)

Observations of the DT's EYCP training and mentoring sessions revealed that all mentors followed broadly the same hub meeting structure and format set out by the DT across hubs and cohorts. Hub meetings generally lasted the expected time of one hour, though three were shorter than expected and two were longer. Practitioners' attendance was generally high, although occasionally Champions arrived late or left early. In one meeting, it was observed that two Champions were unable to attend.

Hub meetings usually began with an overview of the most recent training module and reflections from Champions. Most mentors spent a few minutes going over the format of the hub meeting with Champions prior to beginning. Some mentors used the PowerPoint slides from training (M1 and M7) and summarised the main points in turn, while some used a less structured, more discussion-based approach (M2 and M3). At times (particularly in later stages), some mentors adopted a slightly looser approach which omitted a formal recap section (M2 and M4) in favour of more of a discussion/reflection based approach. All mentors introduced and presented the video exemplar activities and invited discussion and reflection about what had been observed. All used the RNAR (Rehearse, Notice, Analyse, and Refine) reflective forms to structure feedback in observed hub meetings.

A good or high level of engagement was observed in all meetings, with Champions offering reflections and asking questions, although different facilitation approaches were observed between mentors. For example, mentors (M2 and M3) adopted a discussion-based format with few slides, and it was noted that most of the discussion was from the mentors rather than Champions. Mentors (M1 and M7) had to work hard to generate engagement, and there was limited time for in-depth discussion (this was in a meeting which was shorter than expected), and mentors' (M5 and M6) in-person meeting was described as free flowing and unstructured. Discussions appeared to be largely between mentors and Champions, though in a meeting with (M4 and M8), Champions were observed in discussion with each other.

In in-person meetings, Champions were frequently observed in discussions with each other, sharing reflections and offering advice to other Champions. In in-person meetings, it was clear that Champions felt comfortable together as a group. In M2's observed hybrid meeting, which was mostly attended online, it was observed that this format impeded the flow of the session, as online Champions were not always aware that in-person Champions were speaking, so there was some cross-talk.

All mentors were observed using a range of strategies to support engagement including initiating discussion through direct questions, asking follow-up questions, expanding and commenting on Champions' reflections, calling on Champions by name for reflections, offering support and encouragement, and offering reflections from their own practice or settings. Some variation was observed in mentoring style and facilitation approaches across sessions. In a small number of cases, mentors provided more limited opportunities for Champion reflection and discussion, particularly during some online meetings, where mentor talk occasionally dominated interactions. In contrast, observed in-person meetings tended to facilitate more active Champion participation and peer discussion. Variations in mentoring approach appeared to influence the depth of reflective dialogue and engagement during sessions.

Some mentors made some adaptations to the meeting format in accommodating settings' needs. For example, in Cohort 1, two face-to-face meetings adopted a hybrid approach to accommodate Champions unable to attend in person (M2 and M3, and M5 and M6). In Cohort 2, mentor M2 held hybrid meetings for her in-person sessions as several Champions had long travel times. In one face-to-face meeting, only two Champions attended in person (including the host setting) and four joined online.

In the interview, all mentors described following the structure set out by the delivery team which suggested a consistent approach to delivery. Mentors also specifically cited the clear structure and guidance of the training as helping them to ensure their training was delivered consistently across different settings.

> *'Well, it's very structured. I don't see how you can go wrong, really. If you keep to the structure, I mean, it's very, very guided. So if you're sticking to that, then you will have consistency.'* (M4)

DT highlighted some key strategies they used to ensure the training fidelity in the interview listed below:

- All mentors remained in their role across the two cohorts.
- Communication practices stayed the same across different settings.
- Mentor feedback was collected through Google Forms and face-to-face feedback after each session.
- Mentor quality assurance (QA) was conducted through regular review of outputs from meetings and support visits.
- DT also gathered feedback from emails received from Champions/managers.
- DT staff monitor attendance of face-to-face meetings.
- Plan developed with mentors to address any issues around settings not taking part in training.

Barriers to and conditions/facilitators for training delivery

> **RQ1: 1.3. What are the barriers and required facilitators for nominated staff to access and engage with the EYCP PD training and mentoring sessions online and/or face to face (including hub meetings and setting visits)?**

Barriers to training

Time and staffing constraints were the most commonly reported barriers to engaging with EYCP training, highlighted across all stakeholder groups, including practitioners (Figure 9), mentors (Figure 10), and the developers (Figure 11). Participants consistently noted the challenge of releasing staff for training while maintaining adequate staffing ratios. As one setting manager explained, *'covering staff is the main thing… you have to find time to train other staff as well'* (MS39). Similarly, an EYCP Champion noted that while training itself was not overly time-consuming, *'it was just finding the time away from the children'* (CS33). Mentors and the DT also emphasised wider sector challenges, including recruitment and retention issues, which made securing supply cover difficult. These challenges were particularly acute in smaller settings or those with low staff-to-child ratios.

> *'Cover in the nursery is a nightmare because our nurseries, it's not small, but there's CS17 and PS17. And then in the other nursery there's basically me and [the Champion]. So to be honest, we don't. It's not like a twenty-staff nursery. There's hardly any of us so taking anyone out is a logistical nightmare, to be honest.'* (MS17)

Geographical distance and travel requirements also posed barriers, particularly for attendance at in-person hub meetings. Practitioners reported difficulty travelling to more distant locations, with one noting that some meetings were 'too far out... to travel to' (CN25). Mentors similarly highlighted that widely dispersed settings increased travel time, requiring extended staff release and making attendance more difficult.

Two additional practitioner-level barriers were identified. First, changing ingrained practices and mindsets was challenging, particularly in settings with established routines or pedagogical approaches. Second, initial reluctance or anxiety around video reflection was reported, with some practitioners hesitant to engage in this aspect of the training.

Staff were often described as accustomed to existing ways of working, which could limit openness to new approaches. The mentors noticed that setting staff were used to leading interactions with young children and that this behaviour may take time to change.

> *'They're hard-wired to dominate children's conversation, so adults find it very hard to talk with children unless you point it out to them. Even early years practitioners who are very skilled sometimes want to dominate the conversation, and I feel that this project, it takes a lot of time.'* (M6)

For some practitioners, being filmed in the videos was a particular challenge. One Champion mentioned:

> *'I think there's a very practical challenge that anyone going through it has to overcome, which is being filmed. I don't think that usually practitioners are comfortable being filmed and seeing each other on camera... But, on the other hand, the more people do it, the more they get comfortable with it, the more useful it could become.' (*CN37)

In hub meetings, EYCP Champions discussed finding it challenging to implement changes in settings with an established way of working.

> *'It can be difficult implementing changes at her setting because it is a Montessori nursery and they have particular ways of practising.'* (C02H2)

Barriers varied somewhat by cohort (See Figure 9). Time and staffing were the most frequently reported barrier across both cohorts. Cohort 2 reported more barriers related to practice change and organisational factors such as limited engagement with the wider team or 'buddy', lack of leadership support, and the impact of mixed-age groups, while Cohort 1 more frequently highlighted geographical distance and challenges with video reflection, suggesting differences in contextual challenges across cohorts.

The DT also identified additional barriers, including limited access to and confidence in using IT equipment, staff turnover affecting continuity and cascading of training, and physical constraints within settings (for example, space, size, and noise levels). These issues were also observed during training delivery in hub meetings. For example, technical issues with video (sound and picture) were frequently noted by mentors during hub meetings (n = 14; Cohort 1: 5; Cohort 2: 9), and problems with online platforms (for example, Zoom) were observed across both cohorts (n = 8; evenly distributed). Three main barriers were noted by the DT only:

- Access to and skill using IT equipment.
- Staffing changes affecting continuity/cascading.
- Setting environment (space, size, noise).

These challenges/barriers were also observed during the PD training and hub meeting sessions; 'Issues with video (sound and picture)' were frequently observed in hub meetings (n = 14; of which Cohort 1: 5; Cohort 2: 9), whereas for online and hybrid meetings, 'Zoom issues' were observed in both cohorts (n = 8; of which Cohort 1: 4; Cohort 2: 4).

Figure 9: Barriers to training: practitioners' view by cohort

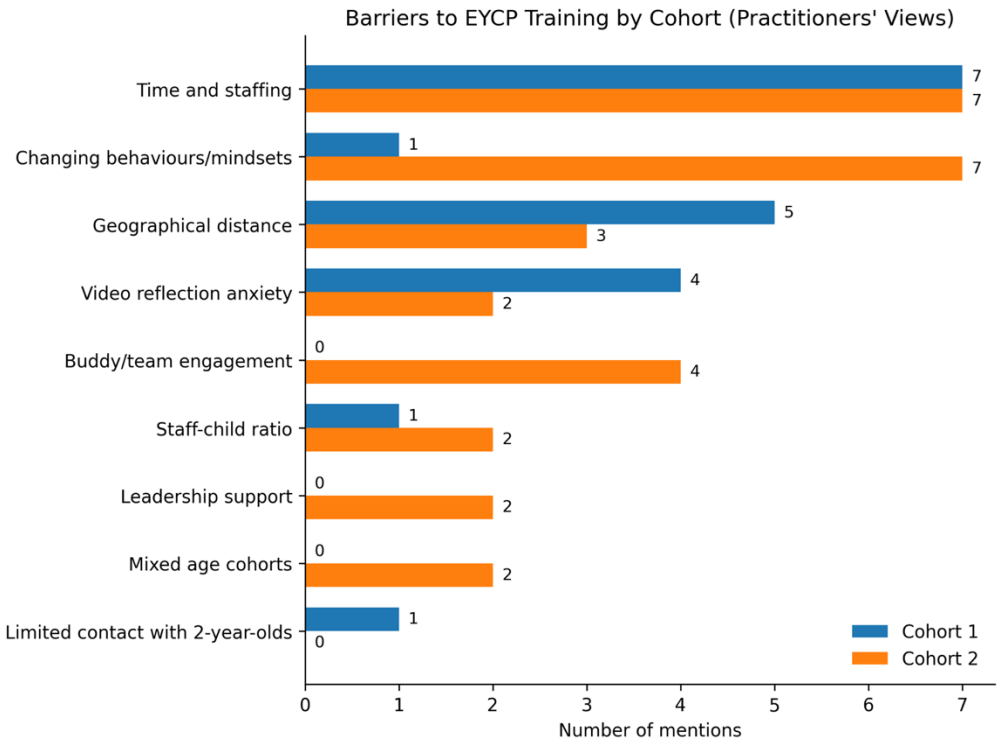


Figure 10: Barriers to EYCP training: mentors' view

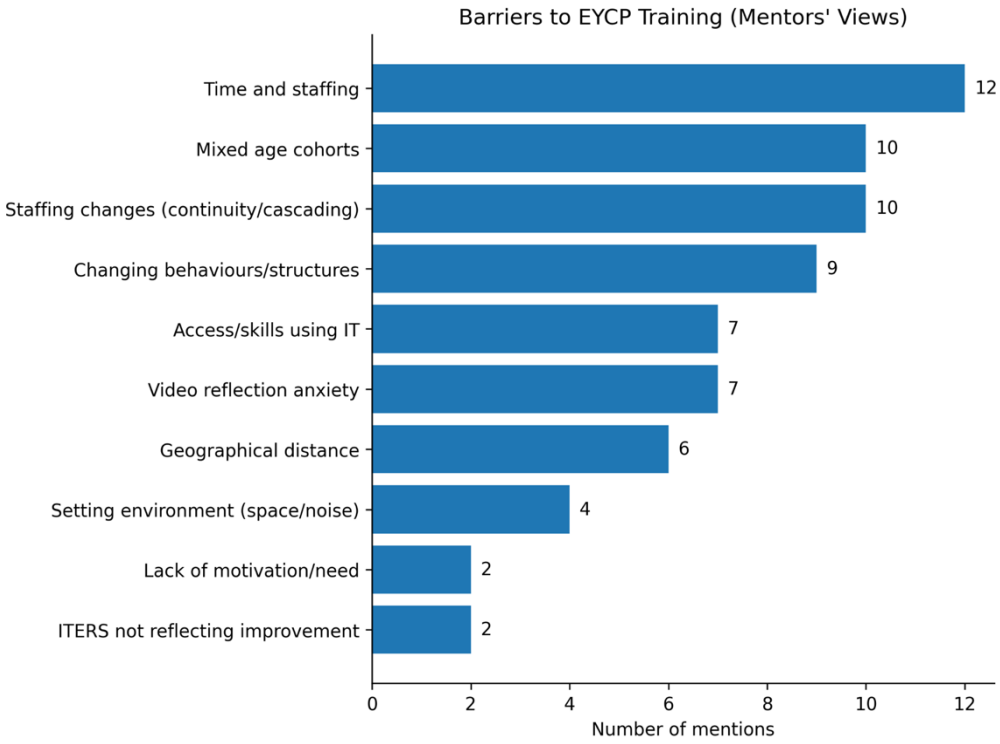
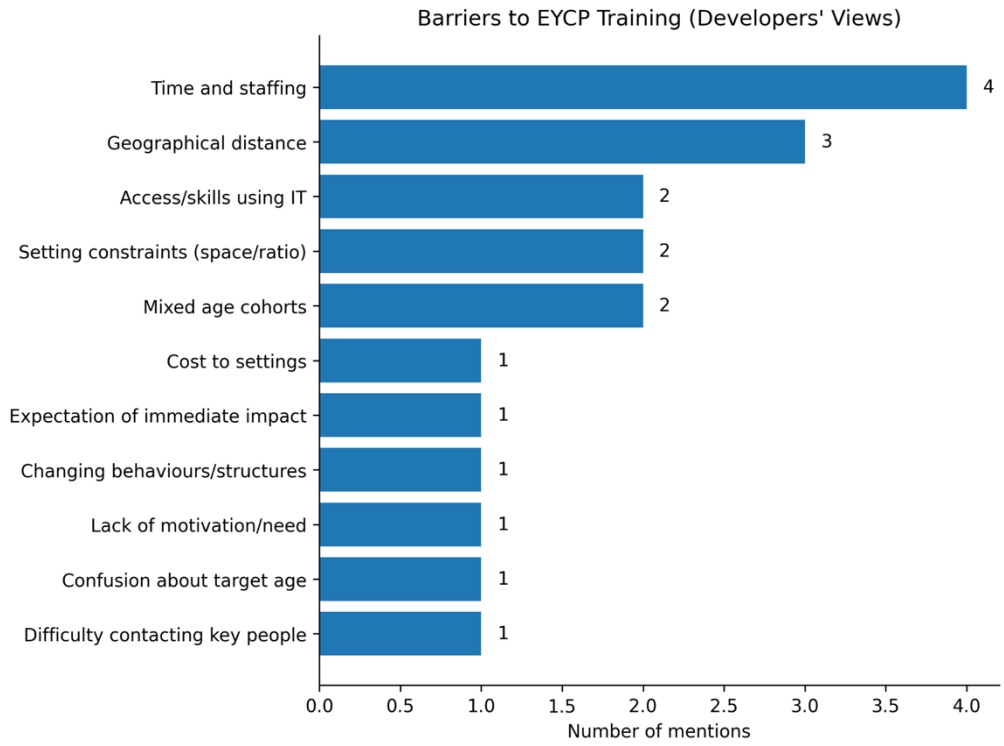


Figure 11: Barriers to EYCP training: developers' view



Facilitators for EYCP training

A range of facilitators supported engagement with EYCP training across stakeholders. EYCP Champions and managers from both cohorts identified setting-level readiness and capacity as key enablers, including willingness to make changes (*n* = 6), the ability to adapt to staffing changes (*n* = 4), and buy-in and support from the wider team (*n* = 4). These findings highlight the importance of organisational culture and team engagement in facilitating implementation.

Some facilitators varied by cohort (See Figure 12). In Cohort 1, practitioners emphasised practical and structural supports, including access to resources, frequent reminders from the Delivery Team (DT), and flexibility in delivery formats, such as the option to attend in-person hub meetings remotely via Zoom. As one EYCP Champion noted:

> *'So accessible, we have everything on the Padlet, the videos we have on the WhatsApp group. And all the time we have a reminder about the meetings, and they send to us on the email as well.'* (CN11)

In Cohort 2, facilitators were more closely related to organisational conditions and practitioner engagement. These included higher staff-to-child ratios, strong leadership support, early observable impacts motivating staff, and ringfencing time for training. Practitioners' enthusiasm, motivation, and willingness to learn were also highlighted as important drivers of engagement.

The Delivery Team (DT) and mentors identified additional facilitators at both the setting and programme level (See Figure 13 and Figure 14). At the setting level, these included advance planning, co-construction of learning opportunities, focusing on small, manageable changes, and *flexibility and resilience in responding to challenges*. At the programme level, facilitators included the multi-strand (hybrid) delivery model, the use of ITERS to structure discussions with settings, the effective use of video-based reflection, and the clarity and consistency of programme messaging. Mentors also highlighted the value of structured tools and collaborative approaches. For example, one mentor noted the usefulness of the ITERS audit:

> *'I found it (initial ITERS audit) really valuable… having a starting point to work with somebody and actually being able to visualise their setting, their existing practices, and the children they're working with.'* (M8)

Finally, developers identified several facilitators related to the mentor workforce. These included mentors' experience in early years and working with two-year-olds, and the collaborative functioning of the mentor team. In particular, it was noted that less experienced mentors benefitted from working alongside more experienced colleagues in Cohort 1, and that continuity of mentors into Cohort 2 supported consistency and quality of delivery. As one developer reflected:

> *'It was good that they all worked together in Cohort 1, as less experienced mentors got to learn from more experienced.'* (DT)

Overall, facilitators to training spanned setting-level readiness, programme design features, and delivery team capacity, highlighting the importance of alignment between organisational context, practitioner engagement, and programme implementation.

Figure 12: Facilitators to training: practitioners' view by cohort

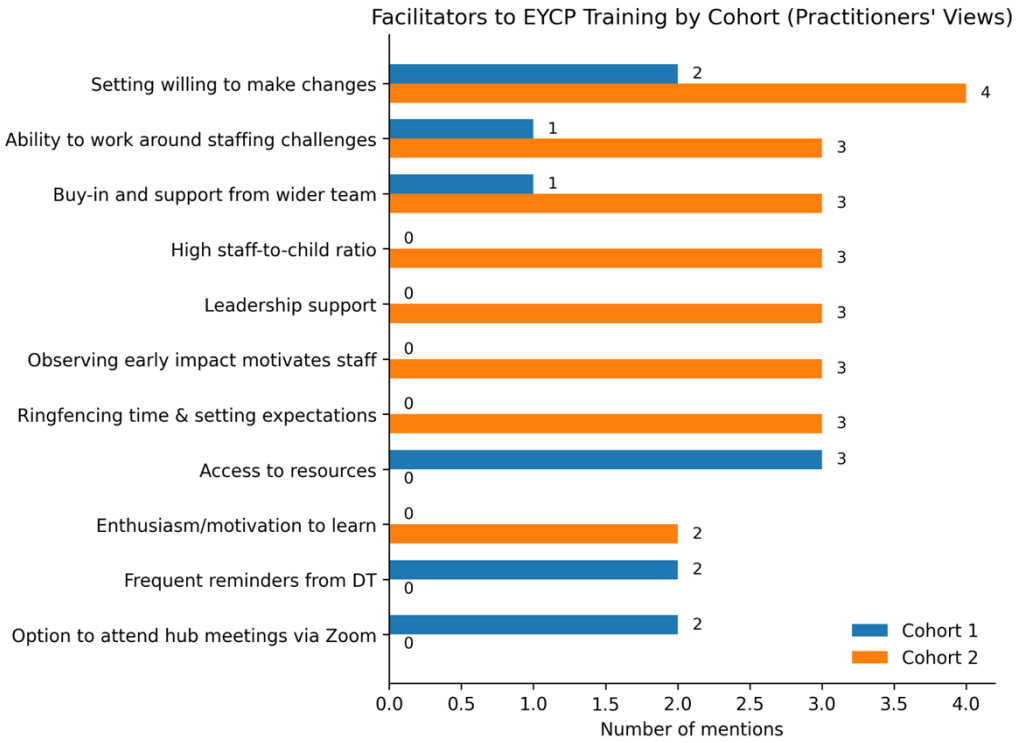


Figure 13: Facilitators to training: mentors' view

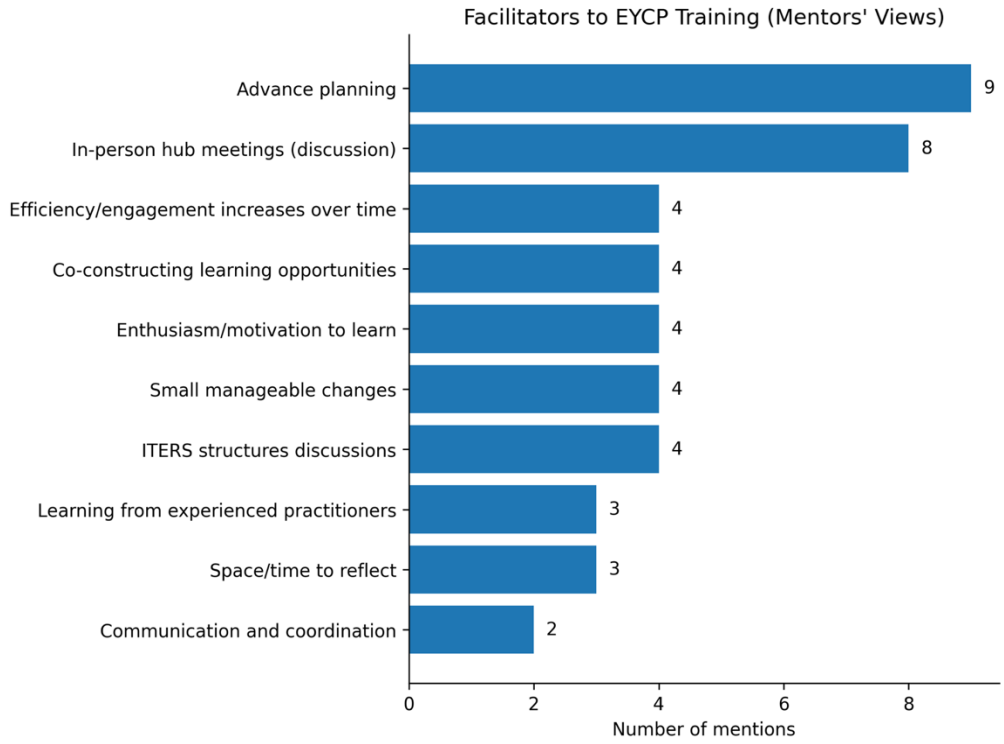
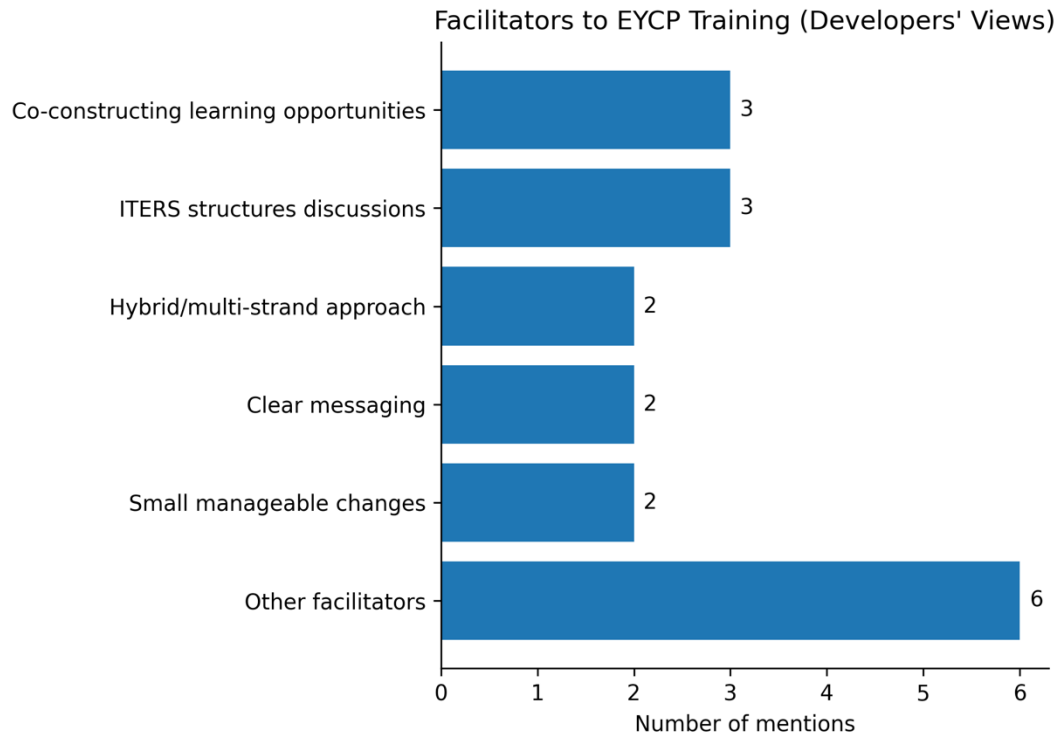


Figure 14: Facilitators to training: developers' view



Fidelity of EYCP programme implementation

This section partially addresses the IPE research question RQ2: 2.1, 2.2, 2.3, 2.4, 2.5, 2.6.

Data used to answer this question has been synthesised from the interview/focus group data with EYCP Champions (*n* = 16), managers (*n* = 11), mentors (*n* = 16), and developers (*n* = 4), hub meeting observations (*n* = 28), and EYCP programme evaluation feedback from the Champions' survey.

Perception of and adherence to the role of EYCP

RQ2: Is the EYCP programme implemented with fidelity in the intervention group in PVI, maintained, and SN settings?

2.1. What are the perceived roles of nominated staff (manager/EYFS lead, EYCP Champion) in the programme?

2.2. Do nominated staff adhere to their roles as requested in the programme?

Champions perceived their role and responsibility in the EYCP programme as twofold: to attend and implement the programme, and to cascade the training to staff within the setting. All Champions (*n* = 16) interviewed discussed their role and responsibility in implementing the learned strategy into classroom practice, whereas many of them (*n* = 11) discussed the EYCP Champions' responsibility to cascade the training to the buddy, room leads, or wider staff within the setting. EYCP Champions from both cohorts discussed that *'It's my responsibility to now pass that on'* (CN17) and *'It's made me sort of really focus on what I'm doing and obviously be more aware of how everyone else is using it.'* (CN25)

A Champion further elaborated on the evolution of the Champion role:

> *'When you first start, it's about embedding your own practice about becoming your own best practitioner... through [attending] your modules that it makes you slow down, it makes you be precise in your actions, and then I think as the year develops, so does your role develop. ...Then I got my buddy, and I was working with them, and then it was rolled out to all five of the staff and the nursery and getting them to film themselves and reflect on what they've done and drip feed, you know, at the end of the day.'* (CS39)

For managers, many of them perceived their role as a facilitator in the EYCP programme, to support EYCP Champions to attend EYCP training and implement the programme in practice. This support included allowing Champions time and space to reflect on, absorb, and discuss the training, and to support the Champions to get other staff on board:

> *'My role was definitely to be able to facilitate it and help CS51A and, you know, implement any changes, also you know, sometimes it's not easy for one person to get all the staff on board. So that was my role.'* (MS51)

One manager stated not having much of a role in the process, explaining, '*We're happy to leave EYCP responsibilities to the Champion*':

> *'To be honest, not much. Apart from giving her [the Champion] the calls. Obviously, [the Champion] does feedback to me. I do come and observe in the nursery. I've seen some lovely practice, but [the Champion]'s very, very experienced.'* (MS17)

One EYCP Champion from another setting reported '*had little support from manager/wider team*' which impacted on cascading and sustained implementation.

In summary, the EYCP Champions have relatively clear perceptions of their role and responsibility in the programme, whereas for managers, they had different expectations or interpretations of their role/responsibility in the programme. This aligns with the EYCP's ToC and Logic model which identifies Champions as key drivers of implementation. However, variation in managers' understanding suggests weaker leadership alignment, which may have constrained whole-setting adoption of EYCP practices.

Cascading EYCP to wider team

RQ2: 2.3. To what extent are the EYCP strategies cascaded to the wider team within the setting?

EYCP Champions were asked in an evaluation form in the endline survey about their general practice in cascading the EYCP programme to other practitioners. Of 33 respondents in the evaluation form, 29 (87.8%) Champions reported that they have been cascading the EYCP training in the setting, of which 14 Champions reported cascading EYCP strategies to the buddy (another practitioner working with the Champion in the classroom), and 15 Champions had cascaded EYCP to the wider staff team beyond their buddy in the classroom.

EYCP Champions shared some examples of using the 'buddy system' in their practice:

> *'My colleague (buddy) who works with me. We are recording each other's practice and reflection on what we can improve using the ShREC approach.'*

EYCP Champions made efforts to cascade the training to the wider team beyond the buddy system:

> *'Every week for 30 mins, I dedicate time to meet with the staff to discuss any observations and to share ideas on how to support children's language development. At least once every two weeks, I record a member of staff interacting with children, then we review it as a team and reflect on the practice and plan next steps.'*

The views from the survey were supported by the setting interview/focus group findings, while all settings interviewed discussed cascading the training to the wider teams to some extent. In the interview/focus group, EYCP Champions discussed the main cascading methods such as:

- use of videos and RNAR forms in the buddy system;
- sharing training resources (for example, Padlet) with other practitioners;
- modelling the ShREC approach in practice;
- introducing interactive book reading; and
- presenting the programme in staff inset day.

During hub meetings, EYCP Champions also discussed cascading relevant EYCP strategies to the wider team in nine hub meeting observations (Cohort 1: 3; Cohort 2: 6). Developers also identified that numerous settings were going beyond cascading to buddies and were implementing whole-setting training:

> *'Lots of groups went beyond the buddy system and were doing whole staff training or parent programmes.'* (Cohort 2 endline)

Mentors also thought that they were seeing the ShREC strategies filtering down to the wider setting: *'I feel like the practice is starting to filter down.'* (M6)

Although cascading of EYCP within settings was not positioned as a core component of the EYCP programme in the Logic Model in this trial (see Figure 1), it was widely adopted in practice, revealing an important emergent mechanism within the implementation of EYCP. However, as cascading was not tightly structured or consistently supported within programme delivery, there was variation in how it was enacted across settings. This variability likely influenced the extent and consistency of practice change at the whole-setting level and may help explain differences in implementation fidelity and outcomes. Given that only around half of the settings were cascading the programme beyond one additional staff member (aside from the EYCP Champion), children's exposure to the intervention within the setting may have been more limited than would otherwise be expected. This may help explain why practitioner outcomes and the mechanisms outlined in the EYCP logic model are generally better supported by the evidence than child-level outcomes, as the effect on child outcomes logically depends on the level of exposure to the intervention. Exposure, in turn, is influenced both by how widely the practices are implemented within the setting and by the amount of time children spend there.

Barriers, facilitators, and other contextual factors for implementation

RQ2: 2.4. What are the barriers and facilitators for implementing the programme in PVI, maintained, and SN settings?

2.5. What other factors explain variation in the fidelity of implementation in the intervention group?

Although EYCP Champions were proactive in implementing and cascading the EYCP programme in their settings through their active engagement in sharing the programme strategies and resources with other practitioners, this process may be hindered by existing challenges. In addition to those barriers for training delivery discussed in section RQ1.1, barriers to implementation, such as lack of staff and time (setting level), established setting structures and routines (setting level), and ingrained (staff and child) behaviours were also revealed by Champions and managers during the interview.

Champions from both cohorts commented on the lack of time and staffing issues for implementing the programme:

> *'I think the challenge is a lack of staff or not having enough time to be with the children more.'* (CN44)

> *'We don't get a lot of time or hardly anything to actually meet together, so that I can maybe go over some stuff or I can like, pass some information on to talk about what I've been going over.'* (CS46)

It was a particular challenge for staff working in the 2-year-old room with a large group size, and the buddy system could put an extra strain on the staffing issue. An EYCP Champion commented:

> *'It's only two of us in the room of 2-year-olds; it's quite hard to sometimes sit down. We do with the interactive book reading, but it's quite hard, but we do it a lot more in our practice.'* (CS17)

One manager acknowledged:

> *'Working with your buddy, that's difficult for staffing, and it was quite difficult to do things in the room, so they would come out of the room, and that's fine, but then that's two members of staff down, so it was just finding the time that worked.'* (MS33B)

Settings with a rigid curriculum structure or an established routine reported difficulties in implementing the programme:

> *'Sometimes we are very bound by a structure. Our nursery has a routine which kids have to go out to play at this time. So the free flow to try and put this [EYCP] in with [was] a bit tricky; we still sometimes find it hard, but we try and make you know there's something needs to go to make this happen.'* (MS35)

This was also echoed by the DT's interview that barriers to implementation include difficulties presented by rigid structures already existing within settings, such as daily routines.

It is worth noting that despite the challenges, most Champions and managers were positive in implementing and cascading the EYCP programme in their settings through their active engagement in sharing the programme strategies and resources with other practitioners.

Regarding facilitators for implementation, the 'Leadership support with setting' was highlighted by Champions, managers, as well as mentors. All mentors discussed the support of the leadership team as being integral to the success of the programme. EYCP Champions and managers commented on providing feedback to their managers regularly regarding mentor support visits and hub meetings.

> *'...we're having those meetings throughout. It's more to kind of get approval for like cascading things; she needs to kind of approve that, or things like the inset day. She needs to know what's happening.'* (CN35)

> *'The first thing is the person who's leading it needs to be on board and understand it. I think that is the absolute main thing.'* (MS51)

Managers also emphasised the essentiality of having their staff with open mindsets for changes, new learning, and development in order to embed EYCP in their settings (MN17, MN09).

> *'It's having the staff be open to you trying things and not having closed mindsets. It did require them to be reflective and sometimes change the things they have done for a long time. We just required the staff to have a positive attitude towards trying things, and luckily, our staff did.'* (MN17)

In addition to these barriers and facilitators discussed above, the setting environment was often discussed as a contextual factor by mentors. In most cases, this referred to the layout of the setting, particularly in the case of private nurseries, which were often in large church halls or converted residential buildings which were noisy or cramped. Often, this was exacerbated by mixed cohorts of children. One mentor explained:

> *'One of the settings that I work in has mixed 2- and 3-year-olds together in a large space. So, you have the double challenge of trying to target the 2-year-olds while managing the 3- and 4-year-olds.'* (M1)

Private vs maintained settings were another commonly discussed topic among mentors. It was generally felt that maintained nurseries had a higher starting point in terms of their usual language practice and the expertise of the practitioners. They also had greater capacity and resources to make changes to their environment and practice. During developer interviews, it was noted that smaller PVI settings may face fewer barriers to implementation compared to other settings, such as School-based nurseries (SNs). While schools generally have more established structures to support the implementation of CPD programmes such as EYCP, larger primary schools were sometimes reported to lack specialist expertise in working with two-year-olds. This was identified as a barrier to effective implementation within these settings.

The Champions' baseline confidence and capabilities were also discussed as a contextual factor. It was felt that less confident/capable practitioners needed more time to implement changes, whereas more capable practitioners who had a higher starting point were able to quickly act on the feedback and lessons from training.

Staff turnover in settings was also noted by mentors as a challenge, as this resulted in changes to the Champion role partway through, or a buddy leaving or moving to another room, which impacted the extent to which Champions could cascade the training to the wider team. One mentor mentioned that:

> *'It was a challenge in terms of the setting started off with one person who had been identified as a Champion who was a 2-year-old room lead. And then, for some personal reason, that person couldn't continue in that role, and they nominated another Champion. She was given time in the 2-year-old room, but it wasn't where she's placed all the time, and so that could be a challenge for the support meetings when you work with the children... she didn't know those children in the morning so well. So we had to make sure that all future ones took place in the afternoon, where she knew the children better.'* (M4)

Although not discussed in Cohort 1, the Cohort 2 interviews referred to the impact of the socio-economic status where these settings were by the DT. Some of the mentors reported having settings in very privileged areas of London which was outside of their previous experience. It was generally felt that children in these settings had a higher language baseline than those in more deprived areas. This variation in starting points may have influenced both the perceived need for the programme and the extent of observable change, with settings serving more disadvantaged communities potentially having greater scope for improvement, while those in more advantaged areas may have demonstrated smaller gains due to already higher baseline levels.

Fidelity of implementation by cohort

2.6. To what extent does the EYCP implementation in Cohort 1 settings differ from the implementation in Cohort 2 settings?

At the programme level, the DT highlighted that programme evaluation forms and hub meetings informed changes made since the start of the programme. They acknowledged that the experience from Cohort 1 provided an opportunity to enhance the programme delivery and to address emerging issues/questions, although basic strategies and structures remained the same.

DT highlighted a few changes to delivery mode and content across cohorts, including:

- increased focus on interactive reading as a core activity;
- increased use of video exemplification in Cohort 2;
- the inclusive practice module was moved to the start;
- removing the behaviour change section and putting it into the training day;
- shifting to RNAR and more support provided in using RNAR in Cohort 2;
- training streamlined for Cohort 2 following feedback;
- updated evidence base for Cohort 2;
- the buddy system is more prominent in Cohort 2; and
- improved communication around support setting leaders have to give the Champions.

All eight mentors from Cohort 1 remained in Cohort 2, but instead of working in pairs in Cohort 1, most mentors in Cohort 2 worked individually, apart from one pair, to support more settings in hub meetings and in setting visits. It was noted that the increase in settings per mentor (Cohort 1: two mentors per hub; Cohort 2: one mentor per hub) has increased workload for mentors in Cohort 2, but on the other side, mentors reported increased familiarity with the training structure, increased confidence and skill facilitating engagement, and increased clarity around what was expected from settings at the outset in Cohort 2.

Whereas mentors in Cohort 2 largely kept the training delivery to the same structure and format as in Cohort 1, three main changes to the training were discussed by mentors. The most salient of these was the increased focus on interactive reading as a core vehicle for the ShREC strategies in the training received by Cohort 2 settings. Mentors felt that this was a positive change, as it provided a focal point for the settings and the activity is naturally conducive to engaging in back-and-forth interactions. Mentors described this as being helpful for less confident Champions who might feel overwhelmed by attempting to immediately use the strategies in everyday practice and interactions. However, although most felt that Champions did begin to expand the use of ShREC strategies into other contexts, others felt that focusing on interactive reading prevented Champions from seeing ShREC as a 'way of being', resulting in it being used as a discrete intervention.

The second change was the baseline ITERS being carried out prior to the launch day training in Cohort 2 (this was done after training in Cohort 1). All mentors who discussed this felt that this was beneficial, as it allowed them to start building relationships with settings prior to training and provided a more valid measure of their usual practice, as in Cohort 1, settings may have already implemented changes resulting from the training on launch day.

Lastly, mentors discussed modifying the hub meetings to include a short tour of the settings hosting the in-person meetings. This change seemed to take place partway through Cohort 1 and was continued into Cohort 2. Mentors perceived this addition as a potential incentive for Champions to attend meetings, as it provided opportunities to observe practice in different settings. While it is unclear whether this change led to increased attendance, it was highly valued by Champions (n = 8), who appreciated the opportunity to engage with other practitioners and visit different settings.

Usual practice

Summary of findings

- Prior to the EYCP programme, practitioners across all settings were already engaging in several language-supportive strategies to a moderate-to-high extent. However, more complex practices—such as multi-turn conversational scaffolding and sustained dialogic exchanges—were less consistently embedded in practice.

- Distinctive features of EYCP implementation were observed in intervention settings including a shift from questioning to commenting, pedagogical transformations with intentional pedagogy and reflective practice, and a focus on interactive reading.
- At baseline, both intervention and control settings had access to a range of language-related CPD, with intervention settings reporting somewhat greater exposure. However, provision was inconsistent, and a notable proportion of practitioners and managers reported no allocated CPD time.
- EYCP was perceived as a unique PD programme: more tailored to the 2-year-olds group, less resource-heavy but also less prescriptive, and more reflective and interaction-focused.

This section addresses the IPE research question RQ3: 3.1, 3.2; RQ4: 4.1, 4.2, 4.3, 4.4.

Data used to answer this question has been synthesised from the practitioner endline survey (*n* = 83), setting interview (*n* = 23), mentors' interview (*n* = 20), developers' interview (*n* = 4), and hub meeting observation (*n* = 28).

Usual practice in teaching language and communication in intervention settings

RQ3: What is the usual practice in language and communication provision like in all nursery settings?

3.1. What are the key observable features of a setting delivering the EYCP (in terms of interactions with children, etc.) in the intervention group?

3.2. To what extent do the EYCP features differ from the usual language and communication practice in the control group?

In the endline survey, EYCP Champions reported on whether they have been cascading EYCP strategies to the wider team, how often they have used the ShREC approach and interactive book reading in their practice in the endline evaluation form.

Of 33 respondents, 29 (88%) Champions reported cascading EYCP strategies in their setting. In terms of the 'use of the ShREC approach', 26 Champions out of 32 (82%) mentioned using it in everyday practice, which suggests a proactive implementation of the ShREC approach in settings. As for the 'use of interactive book reading', 20 Champions out of 32 (63%) reported using it every day, indicating that interactive book reading has also been used in daily practice, albeit not as often as the use of the ShREC approach (see Table 30).

These findings suggest that while both strategies were adopted in practice, the ShREC approach may have been more readily integrated into daily interactions, whereas interactive book reading, although widely used, may require more structured opportunities to be consistently embedded.

Table 30: EYCP Champions' use of interactive book reading and ShREC approach in practice at endline

Use of Interactive Reading	Frequency (*N*)	Percentage (%)	Use of ShREC Approach	Frequency (*N*)	Percentage (%)
Every day	20	63%	Every day	26	82%
More than 3 days a week	6	18%	More than 3 days a week	2	6%
2–3 days a week	4	13%	2–3 days a week	1	3%
1 day a week	1	3%	1 day a week	2	6%
Rarely	0	0%	Rarely	1	3%
Hard to tell	1	3%	Hard to tell	0	0&
Grand Total	32	100%	Grand Total	32	100%

For intervention settings which are implementing/delivering the EYCP programme, some key observable features of the EYCP programme were identified in practice from the interview, including a shift from questioning to commenting, pedagogical transformations with intentional pedagogy and reflective practice, and a focus on interactive reading.

EYCP Champions reported commenting more than asking questions in interactions with children after doing EYCP, although initially using EYCP strategies was hard, as it is going against prior training around questioning. Practitioners were trying to give more time to the children when they speak and to do more small group activities in reading.

> *'We have split our story times, so that we've got smaller groups now, so that we can actually have better conversations with the children. And that they can have more of an interactive story time. And we found the bigger groups' stories was too much. And we've noticed that since doing the training.'* (MN17)

One Champion (CS39) stated that children were not engaged when previously practitioners used lots of questions in their interactions with children, and EYCP has helped practitioners change their teaching to improve engagement (CS17).

The programme has changed the way the Champions approach teaching language development, and the EYCP training encouraged practitioners to slow down and be more precise in their actions with children. EYCP Champions and managers commented on incorporating more language activities into their activities than before, while being more intentional in their interactions with children, around language and communication (CS33).

> *'I don't work directly with the children, but when I walk through the corridor and have conversations with children. I am really thinking carefully about what we have learned through the EYCP. I think, 'What can I do to make an opportunity with it?''* (MS33B)

EYCP Champions noted increased appreciation of the importance of interactive reading strategies and of focusing on children who need support when reading (CS51A, CS17).

> *'It's become very normal, especially like they're not asking questions and commenting rather than asking. That's become very natural and interactive reading we do a lot of... so that the children have more opportunity for conversation and playing and things like that, and less direction and instruction.'* (CN17)

Mentors and the DT also highlighted key observed features in settings which were delivering EYCP programme, including

- focus on intentional pedagogy;
- use of reflective practice;
- focus on interactive reading; and
- increased quality/quantity of interactions with children in practice.

Usual practice across all settings

A total of 83 practitioners responded to the survey (response rate 82.18%) regarding their general practice in teaching language in classroom practice before intervention. It consists of nine items scored on a 0–4 scale with 4 being 'very much' (indicating highest extent of use of indicated strategy) and 0 being 'not at all' (not using the indicated strategy).

Table 31 shows mean scores (range: 2.92 to 3.87) of practitioners' general practice in using those language and communication related teaching strategies in practice. It revealed that all those strategies have been used to somewhat extent by practitioners in their classroom practice, with 'Getting down to children's level and engaging in their choice of activity' being the most commonly used strategy (mean: 3.87), and 'Constructing multi-turn interactions with the children (by using 'Wh' and open questions)' being the least commonly used strategy among those nine strategies.

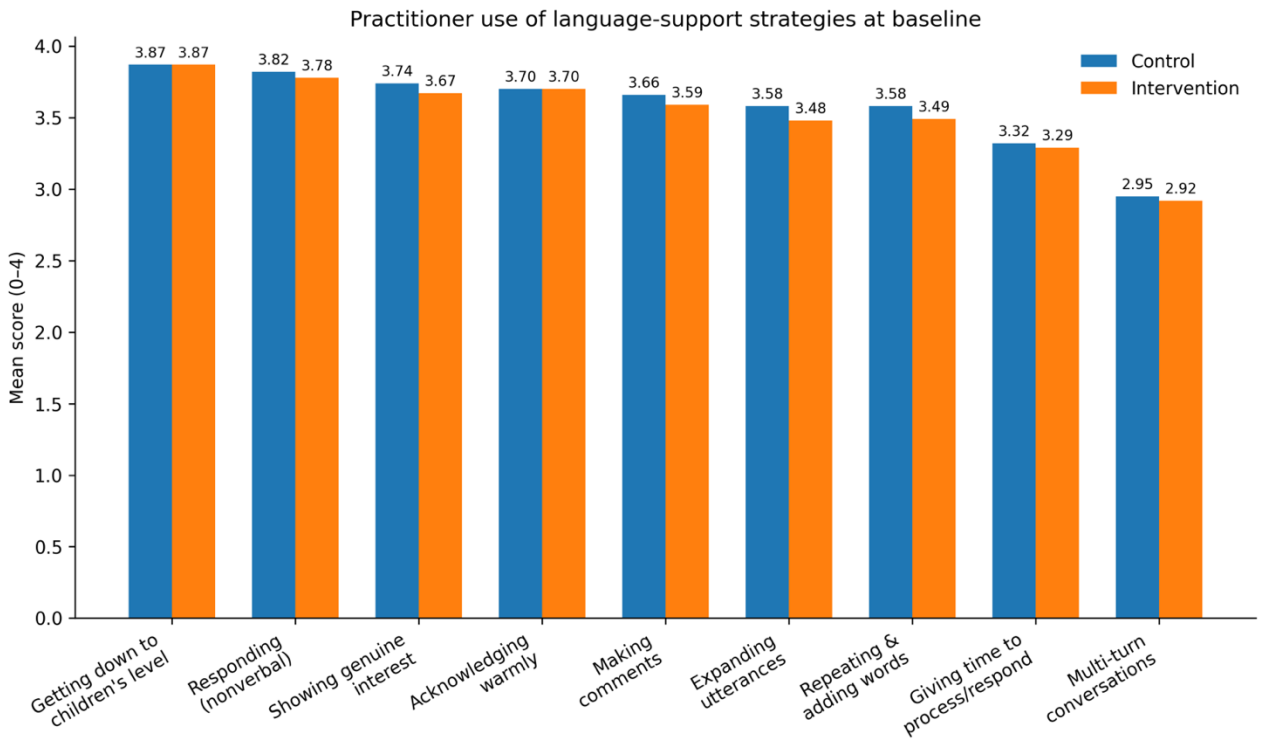
Table 31: Descriptive statistics on practitioners' general practice in using language and communication related teaching strategies in practice at baseline

General practice	*N*	Minimum	Maximum	Mean	Std. Deviation
Showing genuine interest in what young children are focused on to establish a connection.	83	2	4	3.67	.521
Getting down to children's level and engaging in their choice of activity.	83	3	4	3.87	.341
Acknowledging children's verbal and non-verbal communication warmly (for example, using body language).	82	2	4	3.70	.514
Making comments on what the children can see, hear, or feel.	83	2	4	3.59	.606
Responding to children by eye contact, looking expectantly, nodding, smiling.	82	3	4	3.78	.416
Expanding the children's language use (for example, from one word to a few more words).	82	1	4	3.48	.671
Repeating what the children say by adding more words.	82	1	4	3.49	.671
Constructing multi-turn interactions with the children (by using 'Wh' and open questions).	83	0	4	2.92	.872
Giving the children enough time to listen, process, and reply.	83	1	4	3.29	.741
Total general practice score	83	20	38	33.14	4.028

The use of each strategy by intervention condition shows a similar pattern with 'Getting down to children's level and engaging in their choice of activity' and 'Constructing multi-turn interactions with the children (by using 'Wh' and open questions)' being ranked the lowest in scale among practitioners from both intervention and control settings (see Table 31). This implies that before the introduction of the EYCP intervention, both intervention and control settings showed similar patterns of practice to build the communication and language skills of children.

In addition, the other three strategies including 'Responding to children via non-verbal contacts', 'Making comments…' and 'Giving enough time to children to listen, process, and repeat' were ranked with same places in intervention and control settings; four strategies including 'Showing genuine interest in what young children are focused on…', 'Acknowledge children warmly…', 'Expanding children's language use…', and 'Repeating and adding words…' were ranked similarly with only one place difference between intervention and control settings. These findings suggest that while foundational interaction strategies were already well embedded in practice, more complex language-supporting techniques—such as sustained, multi-turn interactions—were less consistently implemented at baseline prior to the EYCP intervention among settings. The ranked order of each strategy in control and intervention settings can be seen in Figure 15 . The close alignment between groups at baseline supports the validity of subsequent comparisons and highlights multi-turn interaction as a key area targeted by the EYCP intervention.

Figure 15: Comparative use of language-support strategies at baseline by intervention condition



Practitioners' responses to all nine strategies in the usual practice scale were combined to represent an overall general practice profile for practitioners. A one-way analysis of variance (ANOVA) was conducted to compare mean differences in general practice measure scores at baseline across cohorts between intervention conditions. The ANOVA results (see Table 32) indicated no statistically significant differences in general practice measure scores among the different conditions (intervention vs control) at baseline within each cohort.

Table 32: One-way ANOVA analysis of usual practice score

Cohort	Condition	Baseline				ES	
		N	Mean	*SD*	*F*	*p*	
Cohort 1	Control	13	31.38	4.87	0.000	0.998	
	Intervention	21	31.38	4.18			
Cohort 2	Control	25	32.48	3.72	2.437	0.125	
	Intervention	24	30.75	4.04			
Cohort 1 and 2	Control	38	32.11	4.12	1.385	0.243	
	Intervention	45	31.04	4.07			

A further independent samples t-test was conducted to compare the general practice scores between Cohort 1 (M = 31.38; SD = 4.38) and Cohort 2 (M = 31.63; SD = 3.94). The t-test results showed that there was no significant difference in baseline

general practice mean scores between Cohort 1 and Cohort 2 (t (81) = -0.272; *p* = 0.786) (see Table 33). This implies that the two cohorts had similar baseline levels of general practice.

Table 33: Independent sample t-test of practitioners' usual practice at baseline

Measure	Cohort	*N*	Mean	Std. Deviation	*t*	*p*
General Practice at baseline	Cohort 1	34	31.38	4.38	-0.272	0.786
	Cohort 2	49	31.63	3.94		

These findings are consistent with the descriptive results presented in Figure 15, which show very similar patterns of strategy use across control and intervention settings. Together, this provides strong evidence of baseline equivalence in practitioners' general language-support practices. From a theory of change perspective, this suggests that any differences observed at endline are unlikely to be attributable to pre-existing differences in practitioner practice. In addition, the baseline pattern—where more advanced strategies such as multi-turn interactions were less frequently used—highlights a clear area of need targeted by the EYCP intervention, particularly through its focus on sustained, high-quality adult-child interactions (for example, ShREC).

Usual language-related PD opportunities

RQ4: **What is the usual PD practice like in all nursery settings?**

4.1. What language-related PD opportunities do staff have in the control group (for example, DfE's Early Years Education Recovery Programme's Online Early Years Child Development Training strand)?

4.2. What are the perceived impacts of these language-related PD programmes on children's language and communication skills in the control group?

In control settings, 36 practitioners (nominated EYCP Champions) responded to questions on continuing professional development (CPD). Of these, 10 (27.8%) reported having no allocated CPD time as part of their role. Among those with allocated CPD time, practitioners reported receiving between 2 and 20 days of CPD per year (mean = 6 days; median = 3 days), indicating considerable variation in access to professional development.

A total of 11 practitioners provided details of language-specific CPD. Reported examples included Language in the Early Years training, Early Talk Boost, in-setting support from speech and language therapists using approaches such as Observe, Wait, Listen (OWL), and training in communication-focused strategies such as Makaton, Watch, Wait, Communicate, and Vocabulary and Early Language Skills.

For managers, of 43 respondents in control settings, 12 (28%) reported no CPD programmes in the previous year at the setting level; 21 managers provided details of previous language CPD at the setting level, such as Social communication, Tales Toolkit, Adult-child Interaction, Early Talk Boost, Watch, Wait, Communicate, Language Enrichment Programme, Speech and Language Therapy (SALT) training, and WellComm.

In intervention settings, 19 out of 44 practitioners (EYCP Champions) (43%) reported they were allocated a certain amount of time for CPD (mean 8.5 days, median 5 days); 19 practitioners provided details of language-related CPD received in the previous year, mostly in-setting CPD from headteachers, also Early Talk Boost, high-quality interactions, and Makaton. From the managers' survey, 21/42 (50%) reported receiving language-related CPD in the previous two years.

Table 34 summarises language PD programmes at the setting level for both intervention and control settings. A good range of language PD programmes were highlighted by settings from both groups, with 'Early Talk Boost' being the most frequently used language PD programme, followed by Speech and Language Therapist support programme and Early Years Professional Development Programme (EYPDP). It appears that intervention settings, compared to control settings, had more language-related PD opportunities before the EYCP intervention. These findings also suggest variability in both access to and type of CPD among practitioners in control settings, with a notable proportion lacking formal CPD provision.

While some practitioners had engaged in language-focused training, this exposure was not consistent across settings. This variability is important in relation to the EYCP Theory of Change, which assumes that structured and sustained professional development is a key mechanism for improving practitioner practice. Differences in baseline CPD provision may therefore have influenced practitioners' readiness to engage with the EYCP programme.

Practitioners perceived these programmes as positively supporting children's language development, communication, social interaction, and engagement. Although direct evidence of impact on children's outcomes was limited, the findings suggest that existing CPD opportunities may have contributed to improvements in practitioner confidence and practice, and may partly explain the modest, non-significant increases in confidence observed among practitioners in control settings.

Table 34: Language-related CPD (setting level at baseline)

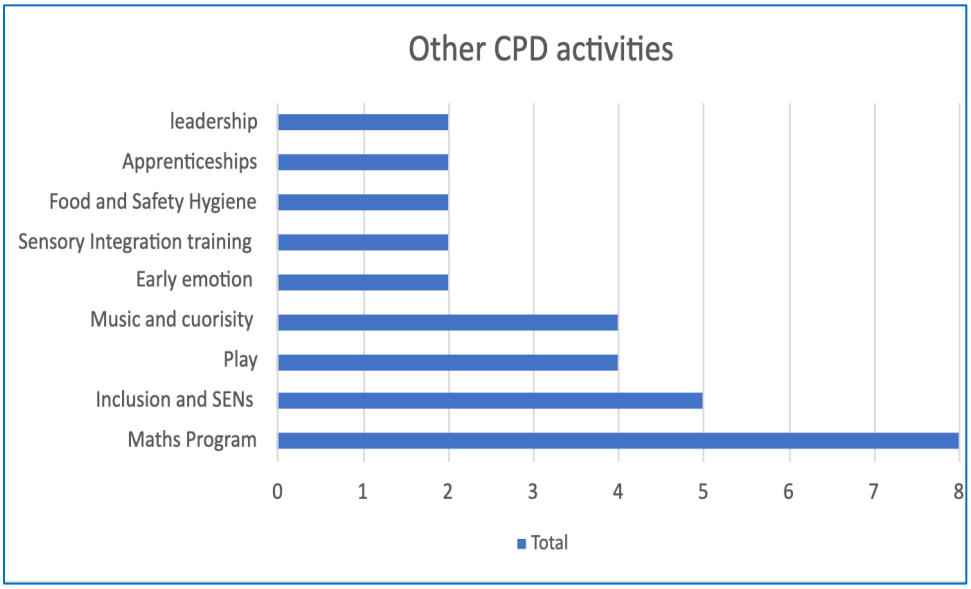
Language-related CPD (setting level)	Total	Intervention	Control
Early Talk Boost	15	9	6
Speech and Language Therapist 'More than Words'	8	3	5
EYPDP (Early Years Professional Development Programme)	6	3	3
Makaton	4	3	1
Merton Early Years	2	1	1
Wellcomm	2	1	1
Language Groups	2	1	1
LASSEY (Language Acquisition Support System Early Years)	3	2	1
Online courses with different companies and programmes	1	0	1
Communication Friendly	2	1	1
Colourful semantics	1	1	0
ACT (Adult-child Interaction training)	1	0	1
NELI (Nuffield Early Language Intervention – Preschool)	1	1	0
Launchpad for Language	1	1	0
Early Years Kick Start	1	1	0
Concept Cat	1	1	0
Making it Real	1	1	0
Chickenshed Project	1	1	0
Overall	53	31	22

Other CPD opportunities

RQ4: 4.4. Are nursery staff participating in any other language-related PD programmes (for example, DfE's Early Years Education Recovery Programme's Online Early Years Child Development Training strand) in the intervention group?

Other CPD opportunities, such as safeguarding, emotional regulation, supporting children with additional needs, maths champion, leadership and management training, were reported in the survey by setting managers/EY leads. Of 26 responses from settings, maths-related PD programmes were mentioned the most, followed by 'inclusion and SENs', 'play' and 'music and curiosity' related PD programmes (see Figure 16). These findings suggest that intervention settings were engaging in a broad range of CPD beyond language-focused training, with a stronger emphasis on areas such as mathematics and inclusion. This may indicate competing priorities within settings, which could influence the attention and time available for implementing language-focused interventions such as EYCP.

Figure 16: Other CPD activities (setting level, provided by manager/EY Lead)



EYCP vs other PD programme: similarities and differences

RQ4: 4.3. To what extent does the EYCP PD programme differ from the usual PD practice in all settings?

During the interview, settings managers and practitioners also discussed previous CPD and provided examples which reflected responses from the survey (for example, Early Talk Boost, NELI, in-house training etc.). Similarities and differences between EYCP and other CPD programmes (language-related) were mentioned during the interview including programme length, targeted (children and staff) group, depth of content, forms of resources as well as staff requirement and feasibility of the programme.

One manager commented that the EYCP programme was similar in duration to other CPD she has engaged but noted that previous training for specific interventions provided more resources for example, a handbook/manual and more guidance about specific books that would be appropriate (for example, for interactive reading).

> *'I thought maybe they (EYCP) would maybe provide more resources. I know obviously a lot of it was about books, but... when we did like [NELI], they give you the textbooks and stuff like that. So I thought there might be a bit more actual resources that were given to us to use... I know that we've got lots of lovely books, but actual more resources probably would have been the main thing. It'd be nice if you had like a like a handbook which was short enough that you could give it to a teacher and she could have a look over it and get a good idea of what the whole course was about.'* (MS17)

A positive difference was that the EYCP programme was tailored to 2-year-olds. *'[EYCP] was the only one suitable for two-year-olds... that's amazing. [having] something for two-year-olds.'* (MS17)

This sentiment was shared by other Champions:

> *'Feels very tailored to us, feels very like bespoke, which is lovely.'* (CS39)

> *'But I think the way it's been delivered, the opportunities that we've had and actually it's really pitched at the right level; I really I've understood everything... I've gone on other training where I've sort of thought. This is really, it's not relevant to my age group, or it's not relevant to my sector. Sometimes I feel like we're lumped in with lots of other. Other sectors... it's nice to have an understanding of it, whereas this is very much about early years. Yeah, and celebrating early years practice, which I think is really important.'* (CS08)

One Champion reported that they had received similar training in the past on interacting with early years children but that the EYCP offered a much more comprehensive experience. Another manager reflected on the difference between the EYCP and another language intervention programme in that it required more staff time away from the classroom, that the EYCP approach was more feasible in terms of staff/time and was more inclusive as you can target more children.

> *'But this programme gave us in depth... She (previous trainer) was just like, here you go read this whereas you guys went into depth with it.'* (CS17)

> *'[The other programme] required a lot of staff time; we took away our reception class staff practically most of the week out of reception. So and obviously when we looked at the ShREC approach, it's an easier approach in terms of timing, but you can impact a lot more children as well.'* (CS39)

Overall, EYCP was perceived as less resource-heavy but more flexible and contextually relevant than other CPD programmes. Its focus on two-year-olds and its practical, embedded approach were key strengths, although some practitioners expressed a need for more structured supporting materials. While the EYCP programme emphasises feasibility, contextual relevance, and integration into everyday practice as key mechanisms for improving practitioner behaviour, the perceived lack of structured resources suggests a potential area for strengthening programme support to enhance consistency of implementation.

Causal mechanism and contextual assumptions

Summary of findings

- Practitioners' survey results indicated that the EYCP programme significantly increased practitioners' confidence in supporting language learning between baseline and endline assessments. There were no statistically significant differences between cohorts in changes in Champion confidence.
- Interview data suggested that the EYCP programme improved practitioner knowledge and intentionality, increased quality and quantity of adult-child interactions, and enhanced child language, engagement, and relational outcomes.
- Stakeholder interviews did not identify substantive differences in perceived impacts across cohorts, despite delivery refinements between them.
- Unintended consequences were reported including changes to settings' structure and routines in some settings with rigid routines.

This section partially addresses the IPE research question RQ5: 5.1–5.9. Data used to answer this question has been synthesised from the interview/focus group data with EYCP Champions (*n* = 16), managers (*n* = 11), mentors (*n* = 16), and developers (*n* = 4), hub meeting observations (*n* = 28), and EYCP Champions' endline survey from intervention settings (*n* = 33), and the intervention setting environment observation ITERS (*n* = 49).

Impacts on practitioners' knowledge, skills, and confidence

RQ5: What are the perceived impacts of the EYCP from different stakeholders (including developers, mentors, setting managers, EYCP Champions, and practitioners)?

5.1. How far does the EYCP PD training and mentoring support impact the knowledge, skills, and confidence of EYCP Champions regarding children's language and communication needs?

5.3. What are the observed impacts of the EYCP upon practitioners' language teaching practice and skills?

Some key observable features of a setting delivering the EYCP in the intervention group were discussed previously in the usual practice section (see IPE RQ 3.1 page 76); EYCP Champions and managers from the intervention group reported practitioners' skills shift from questioning to commenting, pedagogical transformations with focuses on intentional pedagogy, reflective practice, and interactive reading. EYCP Champions also discussed being more confident in teaching language skills to children following the training; this was further supported with evidence from 'the confidence and skills in supporting early language development' measure outcomes.

Practitioners from all settings were asked to self-report their confidence level in supporting children's language and communication learning. Table 35 presents descriptive statistics on practitioners' (EYCP Champions and nominated Champions) confidence level, showing an overall increase in confidence scores from baseline to endline in Cohort 1, Cohort 2, and combined cohort.

Table 35: Descriptive statistics on baseline and endline confidence outcomes by cohort

Cohort	Stage	*N*	Minimum	Maximum	Mean	*SD*
Cohort 1	Baseline	24	12	45	35.00	8.37
Cohort 1	Endline	20	24	48	41.25	6.12
Cohort 2	Baseline	51	19	48	36.33	6.56
Cohort 2	Endline	34	28	48	39.94	5.37
Whole sample	Baseline	75	12	48	35.91	7.16
Whole sample	Endline	54	24	48	40.43	5.62

Figure 17 shows changes in practitioners' confidence from baseline to endline across control and intervention groups. Practitioners from the intervention group reported slightly lower confidence level at the start of the intervention, compared to those from the control group across cohorts. While both groups demonstrated increases in confidence over time, the improvement was more pronounced in the intervention group. A one-way analysis of variance (ANOVA) was conducted to compare mean differences in confidence scores at baseline and endline across Cohort 1, Cohort 2, and the combined Cohort 1 and 2, with group allocation to intervention or control group as the predictor variable, and found no significant difference at baseline and endline confidence outcomes between intervention and control groups (see Table 36).

Next, paired sample t-tests were performed to compare the mean confidence difference between baseline (prior EYCP training) and endline assessments. It appears that after receiving the training, EYCP Champions from the intervention group increased confidence significantly from baseline (M = 34.80; SD = 5.7) to endline (M = 40.05; SD = 5.43), with an effect size of .776 (95% CI: .33, 1.21, $p < 0.001$), which is statistically significant. The nominated Champions from the control group, in the meanwhile, also had a confidence level increase from baseline (mean = 37.44, SD = 8.97) to endline (M = 39.13; SD = 6.71), with an effect size of .256 (95% CI: -.222, .73) which is non-significant (p = .30) (see Table 37).

These findings indicate that while confidence improved slightly in the control group, only the intervention group demonstrated a statistically significant and practically meaningful increase in confidence. The moderate-to-large effect size

indicates substantial gains in confidence among practitioners who received EYCP. However, because the between-group comparison did not reach statistical significance, caution is warranted when attributing these improvements solely to the intervention. While the pattern of results is consistent with EYCP having a positive influence on practitioner confidence, the evidence is insufficient to conclude definitively that the observed gains were caused by the programme rather than by other factors operating over time.

Figure 17: Practitioners' baseline confidence in supporting language learning by intervention condition

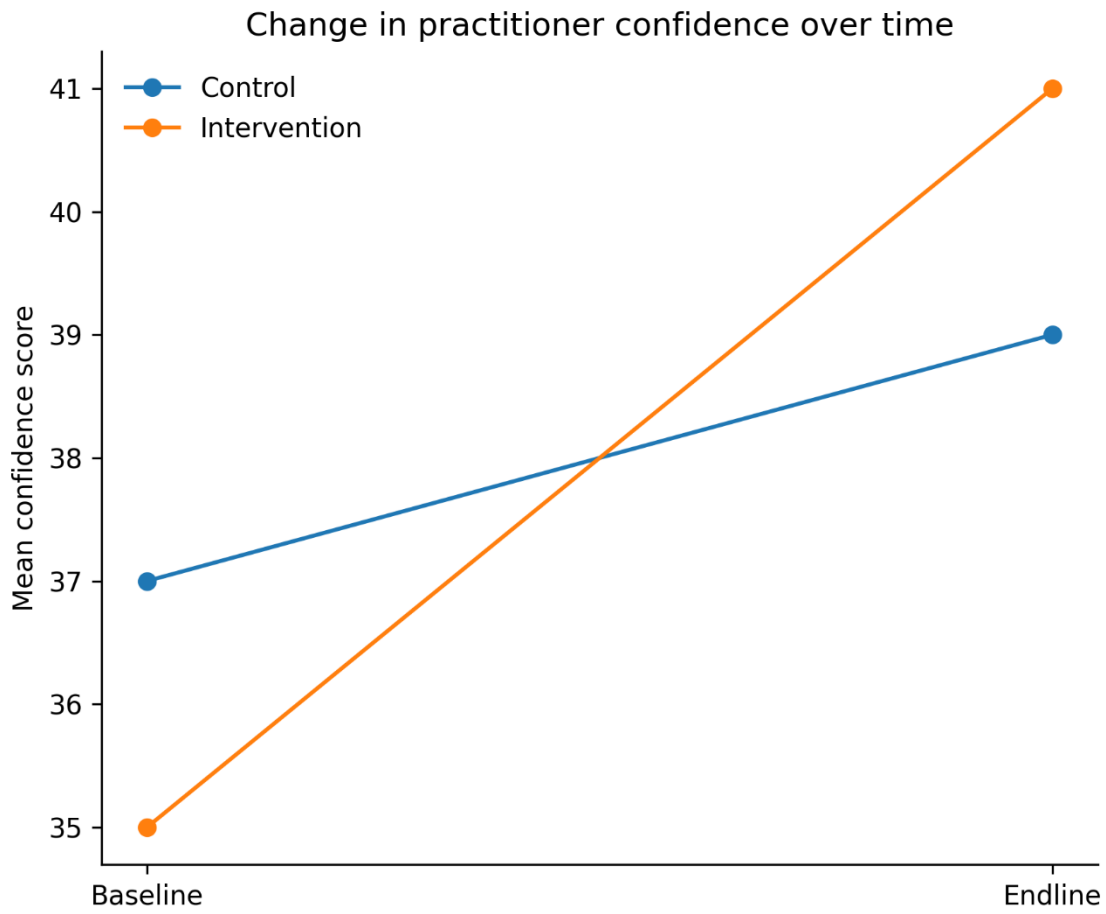


Table 36: One-way ANOVA analyses of practitioners' baseline and endline confidence in supporting language learning in Cohort 1, Cohort 2, and combined cohort

Cohort	**Condition**	**Baseline Confidence**					**Endline Confidence**				
		N	Mean	*SD*	*F*	*p*	*N*	Mean	*SD*	*F*	*p*
Cohort 1	Control	11	35.27	9.9	0.021	0.887	5	36.80	9.36	4.099	0.058
	Intervention	13	34.77	7.1			15	42.73	4.04		
Cohort 2	Control	25	37.28	6.80	1.022	0.317	15	40.67	4.89	0.488	0.490
	Intervention	26	35.42	6.32			19	39.37	5.73		

Whole sample	Control	36	36.67	7.81	0.778	0.381	20	39.70	6.25	0.525	0.472
	Intervention	39	35.21	6.52			34	40.85	5.26		

Table 37: Paired sample tests by intervention condition (1 = control; 2 = intervention) for whole sample

Condition	*N*	Baseline Mean (*SD*)	Endline Mean (*SD*)	Mean Difference (95% CI)	t (df)	*p*-value	Hedges' *g* (95% CI)
Control	16	37.44 (8.97)	39.13 (6.71)	-1.69 [-5.02, 1.64]	-1.08 (15)	.300	0.26 [-0.22, 0.73]
Intervention	25	34.80 (5.78)	40.08 (5.43)	-5.28 [-8.00, -2.56]	-4.01 (24)	<.001	0.78 [0.33, 1.21]

Impacts on the setting's learning environment

RQ5: 5.2. What are the observed impacts of the EYCP upon the setting's learning environment as evidenced by the ITERS-3 audit outcome?

As part of the EYCP programme, the learning environment in intervention settings was assessed by mentors at baseline, prior to training, and at endline, after the EYCP intervention. Of 51 intervention settings, only two withdrawn settings were missing the assessment data on ITERS. Three sets of ITERS data was provided for 49 settings by mentors: ITERS overall scores, ITERS language and book subscale scores, and ITERS interaction subscale scores. Table 38 presents descriptive statistics on three ITERS scores at baseline, endline for each cohort, and combined cohorts. At baseline, settings' ITERS overall scores ranged from 1.50 to 5.80 (mean = 3.76; *SD* = 1.10), suggesting some setting-level learning environment differences across settings, with certain settings below the average score of ITERS (3.5 out of 7). As shown in Table 43, there were increases in ITERS scores from baseline to endline across cohorts and subscales.

Table 38: Descriptive statistics on ITERS overall scores, ITERS language and book subscale scores (SubL&B), and ITERS interaction subscale scores (SubI) at baseline and endline for Cohort 1, Cohort 2, and combined cohort

Cohort	Stage	ITERS	*N*	Minimum	Maximum	Mean	*SD*
1	Baseline	SubL&B	22	1.83	5.50	3.80	0.95
		SubI	22	1.00	7.00	3.82	1.84
		Overall	22	1.71	5.43	3.83	1.01
1	Endline	SubL&B	21	3.20	6.80	5.06	1.12
		SubI	21	3.00	7.00	5.20	1.36
		Overall	21	3.28	6.90	5.08	1.13
2	Baseline	SubL&B	29	1.60	5.80	3.70	1.23
		SubI	29	1.00	6.00	3.69	1.42

		Overall	29	1.50	5.70	3.70	1.20
2	Endline	SubL&B	28	3.83	7.00	5.75	0.91
		SubI	28	4.00	7.00	6.00	0.86
		Overall	28	3.85	7.00	5.79	0.89
1 and 2	Baseline	SubL&B	51	1.60	5.80	3.74	1.11
		SubI	51	1.00	7.00	3.75	1.60
		Overall	51	1.50	5.70	3.76	1.11
1 and 2	Endline	SubL&B	49	3.20	7.00	5.46	1.05
		SubI	49	3.00	7.00	5.65	1.16
		Overall	49	3.28	7.00	5.48	1.05

In Table 39, an independent samples t-test was conducted to compare the ITERS overall scores between Cohort 1 and Cohort 2 at baseline and endline. At baseline, Cohort 1 (M = 3.83; SD = 1.01; N = 22) and Cohort 2 (M = 3.70; SD = 1.20; N = 29) showed no significant difference in their ITERS scores (t = 0.390; p = 0.699). At endline, Cohort 1 (M = 5.08; SD = 1.13; N = 21) and Cohort 2 (M = 5.79; SD = 0.89; N = 28) showed a significant difference, with Cohort 2 scoring higher (t = -2.466; p = 0.017).

Table 39: Independent samples t-tests of ITERS overall scores (baseline and endline) between Cohort 1 and Cohort 2

ITERS overall scores	Cohort	*N*	Mean	Std. Deviation	*t*	*p*
Baseline	Cohort 1	22	3.83	1.01	.390	.699
	Cohort 2	29	3.70	1.20		
Endline	Cohort 1	21	5.08	1.13	-2.466	.017
	Cohort 2	28	5.79	0.89		

*p<.05

Table 40 presents the results of paired-samples t-tests evaluating changes in ITERS scores from baseline to endline for Cohort 1, Cohort 2, and the combined Cohort 1 and 2. It was found that after receiving the EYCP training, intervention settings had made significant improvement on setting learning environment, with effect sizes of 1.540 (95% CI: 1.103, 1.976), 1.284, (95% CI: .885, 1.682), and 1.541 (95% CI: 1.113, 1.969), on ITERS Language and Book subscale, ITERS Interaction subscale, and ITERS overall scale, respectively.

The findings indicate that, following participation in the EYCP programme, intervention settings demonstrated statistically significant improvements in the quality of the learning environment. These large effect sizes indicate substantial improvements in both language-related provision and the quality of practitioner-child interactions, suggesting that the EYCP programme had a strong impact on key aspects of the learning environment. While the results indicate substantial improvements in ITERS scores, they should be interpreted with caution. The assessments were conducted by mentors involved in delivering the EYCP programme, which may introduce potential observer bias. As such, the magnitude of the observed improvements may be somewhat overestimated.

Table 40: Paired-samples t-tests of ITERS scores (ITERS Language and book subscale, ITERS Interaction subscale, and ITERS overall scale) between baseline and endline by cohort

Cohort	ITERS -3 Baseline-Endline	Baseline			Endline			95% CI		t	df	p	ES Hedges'	95% CI	
		N	Mean	*SD*	*N*	Mean	*SD*	Lower	Upper					Lower	Higher
Cohort 1	Language and Book	21	3.80	0.97	21	5.06	1.12	-1.64	-0.89	-6.97	20	< 0.01**	1.46	0.85	2.07
	Interaction	21	3.95	1.77	21	5.19	0.30	-1.93	-0.55	-3.75	20	0.01**	0.79	0.30	1.26
	Overall	21	3.84	1.03	21	5.08	1.13	-1.61	-0.86	-6.86	20	< 0.01**	1.44	0.83	2.04
Cohort 2	Language and Book	28	3.78	1.18	28	5.75	0.90	-2.46	-1.49	-8.36	27	<.001**	1.54	0.99	2.07
	Interaction	28	3.79	1.34	28	6.00	0.86	-2.71	-1.72	-9.11	27	<.001**	1.67	1.10	2.24
	Overall	28	3.78	1.15	28	5.79	0.89	-2.46	-1.56	-9.12	27	<.001**	1.68	1.10	2.24
Cohort 1 and 2	Language and Book	49	3.78	1.09	49	5.46	1.05	-2.00	-1.34	-10.28	48	<.001**	1.54	1.10	1.98
	Interaction	49	3.86	1.53	49	5.65	1.16	-2.21	-1.38	-8.63	48	<.001**	1.28	0.89	1.68
	Overall	49	3.81	1.09	49	2 5.48	1.05	-1.99	-1.36	-10.72	48	<.001**	1.54	1.12	1.97

** p <.001

Interviews with mentors also highlighted settings making changes to structures and routines to facilitate a richer language environment (such as reducing group sizes and transitions, separating two-year-olds from older cohorts, and making changes to how they delivered snack times) and Champions becoming more reflective in their practice.

Impacts on children's language learning and other learning areas

RQ5: 5.4. What are the observed impacts of the EYCP on children's language learning as well as other learning areas in the nursery?

5.7. What are the perceived and observed impacts of the EYCP on disadvantaged children (i.e., with FEE status)? What are the barriers or facilitators that are specific to those disadvantaged children?

Across interviews and hub meeting observations, EYCP Champions and managers consistently reported a range of perceived benefits for children. These included improvements in language development, increased confidence, stronger relationships with practitioners, and positive changes in behaviour.

> *'We noticed a difference in the children almost immediately after starting the programme, and the children really enjoyed doing it with us and it gave them confidence and encouraged their vocab and that encouraged us.'* (MS78B)

> *'I think for the children I've noticed particularly like thinking of specific children where we've implemented interactive book reading as intervention, and as something that happens very regularly that their language has improved throughout the year. So I've noticed it as a really positive kind of impact for them. I think there's more language and conversations in the classroom now because of it.'* (CN35)

Participants frequently described enhanced two-way communication and engagement with children. For example, one manager noted that children were *'spending more time with [staff] and engaging in that two-way conversation'* (MS08), while another highlighted '*huge progress... in terms of their communication'* (CS35). These improvements were often linked to increased child confidence and willingness to interact.

Settings also reported that children appeared more settled and emotionally secure, with some evidence of improved behaviour. One manager described how targeted interactions led to children having more 'quality time', which in turn contributed to calmer behaviour.

> *'CS46 [EYCP Champion] and I had specific children in our rooms that we've been using it with, and their behavioural side has really calmed down... So the children have had that lovely quality time and their behaviour is kind of reduced.'* (MS46)

This suggests that the EYCP strategies may support not only language outcomes but also broader aspects of social and emotional development.

Champions highlighted that the programme benefitted a wide range of children, including those who were quieter, had speech and language delays, or spoke English as an additional language (EAL). For instance, one practitioner described how a previously quiet child became more communicative and better able to form relationships with adults through strategies such as getting down to the child's level and following their lead (CS39). Similarly, improvements in speech were noted among children with limited language following increased use of interactive book reading (CN25C, CN17C, CN35C, CS33).

> *'And with a couple of children whose speech is limited, because from being at nursery and through interactive reading, there is a big difference in their speech.'* (CS33)

> *'I think that particularly for children who are EAL using the books to look at pictures and to give them even that one word vocabulary has really been beneficial for them, and we have a few children.'* (CN35C)

Observational data from hub meetings further supported these findings. Champions shared examples of how applying the ShREC approach led to immediate improvements in children's language use. In one case, a previously quiet child began to use new vocabulary ('red') following a short, responsive interaction with a practitioner.

> *'One Champion described a positive experience with a child following use of the ShREC approach. Child is very quiet and does not use a lot of language at nursery. During nappy change, Champion took opportunity to use strategies by talking about the child's toy car that he brought with him. 'car', 'yes, a lovely red car'; later that day, he showed her the car and said 'red'.'* (C02H3)

In another example, learning key words in a child's home language led to an immediate positive response, highlighting the value of culturally responsive practices: *'Champion two discusses learning a few key words in a child's language and how immediate and positive the impact on the child was when she heard them'* (C02H4).

Impacts on disadvantaged children: barriers and/or facilitators

Champions and managers reported that the ShREC approach was particularly beneficial for children with English as an additional language (EAL), especially during interactive book reading. In contrast to more traditional storytelling approaches—where language comprehension could act as a barrier to engagement—interactive reading enabled children to take a more active role in communication, supporting both engagement and language use.

> *'We have a child who initially she was extremely quiet. English is her first language. It was more to do with being quite unsettled... she's talking way more just by literally getting down to her level and, you know, letting her take the lead and stuff like that, so. That's actually really helped her settle in more, not just her language, but just her being able to make relationships with the adults a lot better.'* (CS39)

Practitioners noted that this shift from passive listening to active participation appeared to increase children's confidence and willingness to communicate. Children who had previously been quiet or disengaged were reported to develop and use more language through repeated, responsive interactions:

'Some of them come with no language… now you can't stop them from talking.' (MS08)

In addition to language development, improvements in behaviour and engagement were also observed among EAL children. These changes were often attributed to increased quality interaction time and more responsive communication: *'We have noticed quite a change… the behaviour has changed a great deal'* (CS51A).

For children with Special Educational Needs and Disabilities (SEND), practitioners reported more varied experiences. Some Champions described adapting the ShREC approach—for example, initially taking the lead to support attention before gradually encouraging child-led interaction: '*I start it off… and then once I see them engage, I take a step back'* (CS35).

However, others reported lower confidence in applying the approach with SEND children and expressed a need for additional guidance: '*Areas we struggled in was children with SEN—maybe more guidance on that*' (CS78).

> '*With the SEN children is quite tough actually, because the ratio I still find that it's still difficult. SEN children, they don't get one to one when they are in 2-year-old setting and also nursery, it takes time for them to get a statement, so giving them the extra time is quite difficult actually. We do have quite a lot of big number of children that attend the nursery. The ratio is one to five now, but one to one children, we do struggle, we do struggle.'* (CN15)

At a setting level, there was also recognition that embedding the approach across all staff could be particularly beneficial for children with SEND. This was seen as reducing reliance on individual practitioners and increasing opportunities for consistent, high-quality interactions:

> *'Having all staff using the ShREC approach… provides the child with a range of communicative experiences.'* (C02H3)

Other contextual factors appear to affect the intervention impacts

> **RQ5: 5.5. What factors appear to affect the intervention impacts (for example, EYCP Champions' qualifications and teaching experience, children's prior language and communication skills, child-level exposure to EYCP, FEE status, and EAL status)?**

Contextual factors that may influence programme implementation at setting level were discussed previously under RQ2.5: Fidelity of Implementation, including leadership support, setting environment, workforce stability, and setting SES status.

At practitioner level, mentors discussed Champions' experience, skills, confidence, and language proficiency as contextual factors that may influence the impact of the EYCP. One mentor (SC) discussed the varying starting points of Champions in her hub and how this affected how quickly they were able to begin implementing changes:

> *'I think my nurseries had quite wildly differing starting points, and I think all were enthusiastic in different ways. So I had a positive group, but some who had a low starting point and kind of got it and were ready to go. I had others with a low starting point who were enthusiastic but were just not quite getting the messages.'* (M8)

Language barriers and confidence were also raised as challenges:

> *'Some of the other challenges have been language. There's been, particularly for one of my mentees, a language barrier around sort of understanding completely and a sort of lack of confidence has been one of the challenges.'* (M1)

> *'So I do think if your own language skills are low, and then you're in a hub with quite articulate, responsive people who've had other professional development and are engaged with it, then that's quite intimidating. I just think her starting point was so low.'* (M8)

At child level, managers and Champions reflected on the impact of having a high proportion of children with additional needs in their settings (with limited staff support), which affected implementation: *'We have so many academic challenges this year (lots of SEN) that it makes it hard to achieve the goals of the project. We are implementing changes, but the lack of SEN support makes it really hard'* (C01H1).

These findings highlight important moderating factors within the EYCP Theory of Change and Logic model, suggesting that the efficacy of the EYCP programme depends not only on the quality of training delivery but also on practitioner capacity and setting context. Variations in baseline skills, confidence, and resource availability may influence how successfully practitioners translate training into improved interactions and learning environments.

Implementation and sustainability

> **RQ5: 5.6 Are settings able to incorporate EYCP strategies into their practice and sustain them in the longer term?**

As discussed in the section reporting on usual practice (RQ3.2), the majority of EYCP Champions reported using the ShREC approach and interactive book reading in their practice in the post-evaluation survey. In interviews, EYCP Champions discussed plans to sustain EYCP in the longer term, including cascading learning to the wider team and/or to older and younger cohorts. These plans included:

- ensuring new staff were trained in the ShREC approach;
- providing regular refresher training throughout the year;
- continuing to work with buddies and the wider team;
- rolling out training and ShREC strategies to staff working with older and younger children;
- creating a setting-wide approach to supporting language development;
- working with parents (for example, one to one or through workshops) to support language development in the home; and
- continuing to embed structured reflection (RNAR) across different areas of practice.

Mentors also commented on settings' plans to continue implementing EYCP beyond the end of the training period. Common themes included embedding structured reflection into routine practice, cascading training to the wider team and older cohorts, and inviting parents into settings to observe ShREC during interactive reading sessions and practise these at home.

In hub meetings, Champions and mentors discussed plans to sustain implementation. Champions' comments were consistent with those reported in setting and mentor interviews (see above). In addition, in Cohort 1, one hub discussed continuing to use the WhatsApp group to stay connected with Champions from other settings, supporting ongoing reflection and shared practice. In Cohort 2, one Champion discussed an intention to strengthen planning for language opportunities, particularly for vulnerable children:

> *'(The Champion) wanted to have a focus on adults—staff and parents. For staff, she wanted to change their planning and to have a new focus on planning specifically for the children that need a bigger slice of the cake.'* (C02H3)

An interview with the DT highlighted strategies to support continuation of EYCP implementation, including establishing an EYCP alumni group/network to formalise meetings beyond the trial period, delivering additional EYCP induction sessions, and exploring future plans for EYCP to be used in different contexts beyond the trial.

Unintended consequences/impacts of the EYCP programme

> **RQ 5: 5.8. What are the unintended consequences/impacts of the EYCP programme in intervention nurseries?**

At the setting level, one unintended consequence of the EYCP programme was the need for some settings to adapt existing structures and routines in order to implement key elements of the intervention. In a small number of cases, this involved modifying daily practices to increase opportunities for high-quality adult-child interactions. For example, settings reported reducing the number of children participating in snack time simultaneously to enable more focused, conversational engagement with adults (N15; S51).

These adaptations were particularly evident in settings with more rigid curriculum structures and fixed daily routines, which required greater adjustments to their organisation and timetabling to accommodate EYCP delivery.

> *'When I went to do the ITERS, the children were well cared for, and they had quite nice relationships. But the children were not talking at all. The setting was very structured: dancing time, book time, circle time, this time, that time. They did lots of transitions by the end they actually absolutely transformed... the practitioner herself said to me about halfway through I realised that the children weren't talking she realised herself. I'd sort of pointed out a few things the timetable is much less rigid.'* (M7)

At the practitioner level, Champions highly valued the opportunities to meet with other practitioners and visit other settings as part of the EYCP training. This positive perception of the EYCP learning community extended beyond the duration of the current trial. In two interviews, mentors reported that Champions in their hubs were planning to continue their peer networks following the end of the training, allowing ongoing opportunities for peer support, discussion, and shared reflection on practice.

At the child level, children from other age groups in mixed-age classrooms were also exposed to the EYCP programme and participated in interactive book reading alongside the targeted two-year-olds during the intervention. Although the presence of mixed-age groups initially presented challenges for practitioners when implementing the ShREC approach and interactive book reading with the target group, strategies were developed to involve older children in supporting and engaging younger children during story time (S08; S78).

> *'The four-year-olds sharing books with the two-year-olds well, and or under 2. So I've seen them in the book corner with the books together, and they're just looking at them together, whereas in the past what we would have observed in the book corner is the older ones sitting the younger ones down.'* (MS08)

Mentors frequently reflected on the positive impact that the training experience had on their own professional development. They described this both in terms of increased capacity to support EYCP Champions and in relation to broader benefits for early years pedagogy and practice within their own settings.

> *'This is my first time doing anything like this. So it's really the benefit to me is, it's really stretching me out of my comfort zone. I'm really excited about. Not that I wasn't excited about my job before. Obviously, I love working with the children. I hadn't come across the ShREC approach before. So it's really changed the way I interact as a practitioner back at my settings.'* (M5)

Impacts of EYCP across cohorts

> **RQ5: 5.9. To what extent does the perceived impact of the programme in Cohort 1 (2023–2024 school year) differ from the programme impact in Cohort 2 (2024–2025 school year)?**

During interviews, managers and Champions reported observed and perceived impacts of the programme. There were no noticeable differences between cohorts in reported impacts at the setting, practitioner, or child levels.

At the setting level, no significant differences were found between Cohort 1 and Cohort 2 in EYCP Champions' change in confidence (as measured by the confidence scale) or in changes to the setting learning environment (as measured by ITERS) at endline.

Interviews with the DT indicated that although there were some modifications to programme delivery between Cohort 1 and Cohort 2 (as discussed in the Fidelity section (RQ2.6)), no perceived differences in programme impacts between the two cohorts were specifically highlighted.

Fidelity, feasibility, and quality assurance of impact evaluation

Summary of findings

- The EYCP programme impact evaluation process followed protocol, with improved operational efficiency in Cohort 2.
- Children's/child-level outcome assessments were delivered with high procedural fidelity within real-world constraints.
- Pragmatic adaptations have been made in terms of setting and child recruitment due to Childcare policy change and recruitment challenges in early years.
- QA processes were robust and positively received.
- The impact evaluation plan was largely feasible within the current trial, with manageable challenges while working with the targeted 2-year-olds group in outcome assessments.
- The evaluation model seems scalable to a larger trial, although targeted refinements around digital systems for recruitment, measurement considerations, and recruitment equity need to be considered.

This section addresses the IPE research question RQ6: 6.1, 6.2, 6.3, 6.4, 6.5.

Data used to answer this question has been synthesised from the ET interview (*n* = 18) and observation of assessors' fieldwork in the QA process (*n* = 38).

Settings, practitioners, and children recruitment

RQ6: To what extent does the EYCP programme impact evaluation process adhere to the plan?

6.1. Does the nursery setting recruitment process adhere to the recruitment strategy?

6.2. Do PVI, maintained, and SN settings respond similarly or differently to the recruitment strategy?

6.3. Does the children and family recruitment process adhere to the recruitment plan?

6.5. To what extent does the setting and children recruitment strategy used for Cohort 1 differ from the strategy used for Cohort 2?

Recruitment of settings and practitioners was led by the DT, supported by the ET, and closely followed the pre-specified recruitment strategy (see Evaluation Protocol) for both cohorts. All 102 recruited settings met the eligibility criteria: 44 from Cohort 1 and 58 from Cohort 2. Overall, 61 settings (59.8%) were private, voluntary, and independent (PVI), and 41 (40.2%) were school-based or maintained settings. DT reported that there had been insufficient PVI representation in Cohort 1 and therefore sought to redress this imbalance with a higher proportion of PVI settings recruited in Cohort 2. Maintained and SN settings responded similarly to the recruitment strategy.

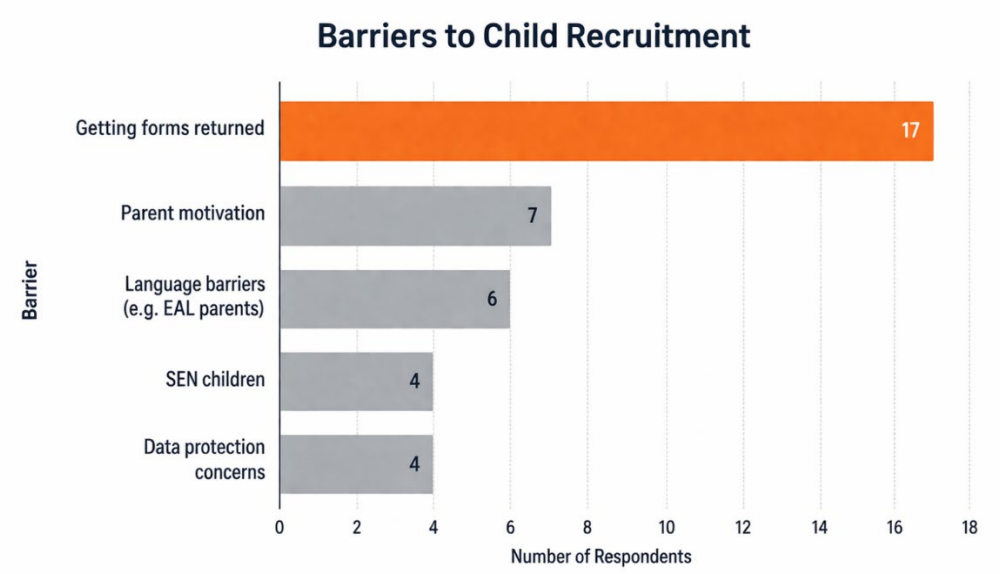
The recruitment process implemented by the DT in Cohort 2 was considered more efficient than in Cohort 1. This improved efficiency was attributed to the addition of a second staff member (increasing the team from one to two), the establishment of more formalised routines for contacting settings, and the introduction of automated processes for signing the Memorandum of Understanding (MoU). Additional recruitment strategies introduced in Cohort 2 included the use of postcards and in-person showcase events held during the summer holidays. DT also reported making greater use of local authority contacts. Despite these enhanced efforts, the DT noted that securing firm commitments from settings to participate in the programme remained challenging.

There were also demographic differences between cohorts. Cohort 2 included a higher proportion of more affluent settings. Following changes to childcare policy implemented in September 2024, funded early education (FEE) places for two-year-olds were no longer an eligibility criterion for Cohort 2. As a result, fewer settings in Cohort 2 were located in disadvantaged

areas (IDACI quintiles 1–3): 27 of 44 settings (61%) in Cohort 1 compared with 12 of 58 settings (21%) in Cohort 2. Additionally, recruitment areas were expanded in Cohort 2 to include South London boroughs in order to meet recruitment targets. However, this geographical expansion was identified as a barrier to practitioner attendance at hub meetings due to increased travel distances.

For child recruitment, the ET led the process, with support from participating settings, which distributed both digital and paper consent forms to parents and carers. Overall, fewer than one-third of parents/carers (28%) provided consent online. In Cohort 1, 16% (61/392) used the online form, compared with 37% (187/502) in Cohort 2. Settings reported that collecting completed paper consent forms from parents was one of the main challenges in child recruitment (see Figure 18). Given the additional time and administrative burden placed on settings, ET reflected that online consent procedures appeared to be both more efficient and more cost-effective.

Figure 18: Difficulties with child recruitment for settings



Administration of children outcomes assessment

6.4. Do baseline and post-test administrators (independent researchers) effectively and appropriately deliver the assessment tasks?

Assessors from both cohorts were interviewed at the end of impact evaluation (*n* = 26), and analysis of assessors' interview data identified five major themes relating to the assessment of two-year-old children using the researcher-administered BPVS and WellComm measures.

Setting assessment environment

Assessors reported conducting assessments across a range of setting environments. Some sessions took place in quiet spaces, while others were affected by background noise or interruptions from other children. Despite these variations, assessments were generally completed successfully, often with the support of a setting practitioner accompanying the child.

> *'Whenever I got to the nurseries, the nurseries were really helpful. All the people were really nice and helpful, and they always made sure I had a little space to go into.'* (A2B06)

> *'The practitioners, the other workers normally did a good job of trying to at least distract the other kids and keep them away from where we worked.'* (A2B08)

The findings highlight the importance of assessor flexibility in adapting to different environmental conditions. While quieter spaces were preferable, assessors demonstrated the ability to manage distractions and maintain assessment fidelity within real-world early years contexts.

> *'We moved to a corridor area with some parallel play nearby, which normalised what we were doing. Where children felt shyer or had just joined and were anxious, sometimes the Champion got another one to one worker they knew would work better with that child.'* (A2B10)

Children's language background and engagement

Assessors emphasised the importance of building rapport with children prior to commencing formal assessment tasks. Most children demonstrated increased engagement when assessors spent time interacting informally beforehand, often in the presence of a familiar room practitioner. Strategies such as playing with toys together and offering small incentives (for example, stickers) before, during, and after the assessment were described as effective in fostering engagement.

This theme was particularly salient for children with SEND or those who were non-verbal.

> *'I think the children learning EAL, I haven't noticed any difference. I don't look at that child and think that this is for this reason. I think there's a marked difference in terms of how some of the children with SEND are engaging because they're either not or they can't sit. You know, they can't be still or they're not looking or they just are finding it difficult to engage in and follow those instructions, and yeah, so I found if I haven't been able to assess children often they've had special educational needs or disabilities.'* (A1E01)

Assessors noted the need to be sensitive to children's fluctuating engagement levels and recognised that some children might be unwilling to participate at certain times. In such cases, conducting a return assessment on a different day or at a different time was considered necessary.

Overall, the findings underscore the central role of rapport-building and contextual awareness in assessing very young children's language skills.

The dilemma of practitioner's assistance

Assessors reported concerns regarding practitioner involvement during assessments. In some cases, practitioners appeared not to consistently acknowledge children's background characteristics (for example, EAL status). Occasionally, some practitioners provided hints or prompts that could influence children's responses, thereby affecting the authenticity and standardisation of the assessment. One assessor shared an example:

> *'So yesterday, it was like a really different environment. The staff kept saying like, oh, they don't know that word. Can you, like, phrase it differently. I'm like, no, I can't, unfortunately.'* (A2E11)

To address this, assessors described proactively communicating with practitioners to clarify assessment protocols. In particular, they emphasised the importance of children responding independently—for example, by pointing to BPVS images without guidance.

For example, assessors were asked to communicate the message with practitioners upon their arrival:

> *'When I went into settings, I'd usually ask if it's possible to have a practitioner to sit with them. But I'd say, like obviously, please don't give anything away or say anything like, just make the child feel comfortable. So they were just observing; they weren't interfering with the assessment at all.'* (A1E03)

It is noteworthy that overall settings were described as helpful and supportive of assessors with their presence during assessments perceived to enhance children's engagement.

> *'If there's been a lack of cooperation from the children, [the practitioners] they've helped—because obviously the kids know the workers a lot better. They've helped to coerce without giving anything away, and helped to get the kids to speak and engage in the assessment.'* (A2B06)

> *'They're the adult the children are most familiar with, and that's the person if they're most comfortable with, they're most likely to help them along and make sure they engage.'* (A1B10)

This theme highlights the tension between maintaining standardised assessment conditions and working collaboratively within early years settings. While practitioner involvement introduced a potential source of bias, assessors actively managed this risk by clearly communicating assessment protocols and intervening when necessary. As such, although instances of prompting were observed, these were limited and unlikely to have systematically affected the validity of the assessments.

Measurement issues

Assessors identified several measurement-related challenges. Assessors observed that some children demonstrated vocabulary knowledge aligned with higher BPVS sets but were unable to sustain attention long enough to complete them. As a result, children may have obtained lower scores than their underlying language abilities would suggest. Although only reported in a small number of cases by assessors (Cohort 1: 3; Cohort 2: 4), this raises questions about the extent to which the measures fully capture two-year-olds' communicative competence, particularly following the implementation of the EYCP ShREC approach.

One assessor shared an example:

> *'I think for me, most of the time it was absolutely fine; the one time it was a bit problematic, because when the child I found was actually really capable, and they were progressing up through the set. But they didn't have the attention to match their ability almost. So you could tell that they were, you know, they might got set 5 or 6, cause I had a couple; it did get 7, but by the time they got to set 6 or 7, you could tell they're starting to look around the room and start looking at toys.'* (A2E14)

Attention span was also identified as influencing performance on WellComm, which requires more verbal responses compared to the picture-pointing format of BPVS. Additionally, the text-heavy format of WellComm was perceived as potentially overwhelming for some children. Assessors commonly reported that children tended to score higher on BPVS than on WellComm. This difference suggests the need to consider how assessment design, attention demands, and response format may influence performance outcomes, particularly in very young children.

Across both timepoints, the BPVS was consistently viewed as easier to administer than WellComm. Assessors reported that children generally enjoyed the BPVS, and no major issues were identified. However, at endline, some assessors observed that children occasionally became bored and began guessing responses in order to finish quickly.

WellComm was reported to be more challenging to administer, particularly at certain items (for example, the 'chair' question, see quote below), which were difficult for both children and assessors to interpret and score consistently, but also appeared to disrupt the assessment flow with young children. One assessor gave an example:

> *'Like with the WellComm, like sometimes when a child's like in the flow of doing the questions. And then you, it's like the chair question [which asks the child, stand up and move]. It can kind of mess the flow up a bit …'* (A1E03)

It was also noted that administering two assessments consecutively could be confusing for children, and transitioning from the picture-based BPVS to the more varied activities in WellComm could be difficult. At baseline, assessment duration ranged widely from 5 to 20 minutes, with distracted children taking up to 40 minutes. By endline, assessments more typically lasted 20–30 minutes per child, suggesting greater consistency in delivery.

Quality assurance and assessor support

Assessors spoke positively about QA and paired assessment experiences, noting that constructive peer feedback enhanced their confidence, procedural understanding, and scoring accuracy. These processes contributed to a greater sense of security when delivering assessments independently.

The findings from assessors' interviews highlight both the practical and methodological complexities of assessing language in two-year-old children. While assessors demonstrated adaptability and professional judgement in navigating

environmental and engagement challenges, the data also raise important considerations regarding standardisation, practitioner involvement, and the developmental appropriateness of assessment tools. These findings highlight a potential tension between children's underlying language competence and their ability to demonstrate this competence within structured assessment contexts. In particular, performance may be influenced not only by language ability but also by attentional capacity, task demands, and response format.

However, while these factors represent potential sources of measurement error, they are unlikely to have systematically compromised the validity of the assessments. The observed challenges appeared to be situational and non-systematic, reflecting developmental characteristics typical of two-year-old children rather than consistent flaws in test administration. Instead, the findings suggest that the measures may be sensitive to attentional and behavioural variability, which should be considered when interpreting scores.

In sum, these results underscore the importance of considering developmental appropriateness and assessment design when evaluating language outcomes in very young children, while supporting the continued use of these measures with appropriate caution in interpretation.

RQ7: What is the feasibility of the impact evaluation plan in this efficacy trial and is it feasible to be scaled up?

This section addresses the IPE research question RQ7: 7.1, 7.2, 7.3, and 7.4.

Data used to answer this question has been synthesised from the ET interview (*n* = 18) and observation of assessors' fieldwork in QA process (*n* = 38).

Parents/carers administered language measurements

7.1. Are parents/carers administered language measures acceptable for parents/carers and are they feasible to use in a large-scale trial in the future?

Two parents/carers administered measurements were used in this trial: the ELIM measure at baseline and the CDI-III measure at endline, both in physical and online forms. At baseline, the ELIM had a high response rate (863 non-missing, 33 missing, 3.7% missing rate), largely due to the administrative arrangement to ask parents/carers to fill out the ELIM forms together with sign-up of consent forms. At endline, the CDI-III had a lower response rate (634 non-missing, 262 missing, 29.2% missing rate) while the parents/carers surveys were distributed at the end of the intervention. As discussed before, settings reported difficulties in getting parents' forms back (see Figure 18), and the long form of CDI-III (100 words) may further prevent parents from completing and returning the forms. Taking all into consideration, those parents/carers who administered language measurements appear to be feasible to use in a scale-up trial, but administrative arrangements need to be in place to ensure a higher return rate such as using a shortened form, motivating participation through incentivisation, and providing assessment supports for parents/carers with difficulties in filling and returning forms.

Working with 2-year-olds group

7.3. What challenges/opportunities has the age of the target population (two to three years old) had for delivery of the evaluation?

Analysis of assessors' interview data and the fieldwork observation revealed a few challenges around working with the targeted 2-year-olds group as part of an RCT trial, including engagement challenges, environmental challenges, and measurement challenges.

Whereas most assessors delivered assessments with children in varied setting environments, meaning some were quiet while some had disturbances with other children, the assessors managed to complete the test properly. However, many of them noted that achieving engagement with assessments was difficult, especially at baseline assessment when most children were only two years old (A2B01, A2B09, A2B06). Barriers to engagement were summarised as shyness, feeling intimidated, potential SEN or EAL status, and those children who are more able and went further in the book sometimes then get bored, impacting engagement.

Various strategies were observed in the fieldwork and found to be useful to achieve engagement. These included: the use of toys (A2B06, A2B10) (although this may also be a distraction depending on the child), having children around to help make the child feel at ease (although this could be counterproductive when they tried to answer questions themselves) (A2B08), stickers (A2B01, A2B09, A2B06), and making assessments feel like a game (A2B09). Giving children regular breaks was important (alongside going through pictures as quickly as possible to prevent them getting bored (A2B09). Assessors also reported that the age of children made a difference with engagement, and those children closer to three were much easier to work with in the assessment task (A2B06, A2B08).

Although children with SENs were not eligible for the impact evaluation due to measurement constraints, it was difficult to capture children's SEN status at baseline assessment at age two, as those children were either not diagnosed, or parents/carers and/or setting staff were not aware of children's SEN situations. Finally, identifying the most appropriate language measures for two-year-olds appeared to be challenging. Whereas the BPVS scale was considered to be appropriate and easy to use, the assessment environment restraints together with the children's development status at age two presented further challenges for researchers to conduct the baseline assessment, which led to a relatively high rate of missing data on BPVS measures at baseline.

These findings highlight the practical challenges of conducting standardised assessments with very young children in real-world early years settings. In particular, variability in engagement, attention, and environmental conditions represents a potential source of measurement error. However, these challenges appeared to be situational and developmentally typical, rather than systematic threats to the integrity of the assessment process. Assessors were generally able to adapt their approach and complete assessments as intended, suggesting that the procedures are feasible within an RCT context.

At the same time, the findings indicate that assessment outcomes may be influenced by factors such as attention span, task design, and setting conditions. As such, results should be interpreted with appropriate caution, recognising that scores may partly reflect children's ability to engage with the assessment as well as their underlying language competence. Overall, the evidence supports the feasibility of conducting these assessments at scale, while underscoring the importance of developmentally appropriate administration and careful interpretation of findings.

Trial scale-up and further considerations

7.2. Are the setting and children recruitment strategies used in the current trial feasible to use in a large-scale trial in the future?

7.4. What factors may need to be considered in scaling up the intervention to deliver in more nurseries?

The DT stated that they know of settings which are keen to continue beyond the trial period. An EYCP alumni group has emerged to formalise meetings beyond the trial period. It was also reported that groups initiated their own staff or parent training programmes containing elements/core components from the EYCP training content shared with practitioners.

When contemplating a larger EYCP trial, the DT stated that they believe the setting recruitment strategy to be scalable. This would be achieved by possibly looking at other technological enhancements such as the development of a coaching app to support mentors while managing hubs in a large trial. Also, if scaling up, they may need to think about improved systems to manage setting records. Recruitment outside of London could draw on the Research School Network—this would allow the use of existing skills/knowledge/experience. Also, strategic networks are important to scaling, as this allows the building of a community of expertise. The building of relationships is identified as important in breaking down some of the barriers to successful deployment. There are also plans to use EYCP in different contexts beyond the trial such as the use of the EYCP programme for different age groups.

These findings suggest strong potential for sustainability and scalability of the EYCP model. Continued engagement from settings and the emergence of alumni networks indicate a degree of programme ownership and embedding beyond the trial context. The identification of scalable recruitment strategies, alongside the use of existing networks and technological supports, further strengthens the feasibility of wider implementation. However, successful scale-up will likely depend on maintaining relational infrastructure and ensuring adequate systems for coordination and quality assurance.

Good/best practice

RQ9: What are the key features of the EYCP 'good/best practice' settings in the current trial?

Overview of characteristics of the nominated good practice settings

Three Cohort 1 and four Cohort 2 good practice settings were nominated by mentors. The majority (*n* = 4) were maintained nurseries attached to infant or primary schools, two were private for-profit settings, and one was a voluntary nursery. Setting size based on number of nursery places ranged from 26 to 52. At the start of training, good practice settings had between 8 and 47 two-year-olds on the roll. Table 41 describes the sample size and characteristics of the nominated good practice settings, as well as reasons for nomination by the DT.

Table 41: Characteristics of the nominated good practice setting

Cohort	Setting ID	Setting Type	Setting size	Reasons for nomination (Delivery team)
1	N06	Private for-profit nursery	Small	Champion demonstrated high levels of reflection during hub meetings and has made a number of positive changes at the setting to routines and practice.
1	N09	Maintained infant or primary school	Medium	High levels of engagement and enthusiasm. Strong evidence of implementation and cascading to wider team and parents.
1	N17	Maintained infant or primary school	Large	Champion described as keen, proactive, and reflective. She was well supported by the EYFS lead, and the setting has made positive changes to implement the training.
2	S17	Maintained infant or primary school	Small	Champion was very motivated and internalised the key learning from the training. She immediately acted on feedback from ITERS making changes to the setting environment. She has successfully cascaded the training to her Buddy.
2	S54	Maintained infant or primary school	Large	Champion has effectively cascaded the training to the wider setting including agency staff. A whole setting wide shift towards child centred/directed interactions and providing opportunities for communication observed by mentor.
2	S58	Voluntary	Small	Champion and manager have engaged fully with the training. The Champion is thoughtful and enthusiastic and has maximised language learning opportunities in the setting. Champion has grown in confidence and modelled ShREC for colleagues.
2	S78	Private for-profit nursery	Medium	Champion and manager attended all training and are fully invested in the EYCP. Champion is enthusiastic, reflective, and constantly challenges herself to improve. ShREC is fully embedded into the setting.

Overview of the data source

Of the seven good practice settings, only two (N09 and S17) provided complete survey data (for example, manager and Champion completed both baseline and endline surveys). All seven Champions completed baseline surveys, but only three completed endline surveys (N09, S17, and S58). Two managers did not provide any survey data (N06 and S54), four managers completed only baseline surveys (N09, N17, S17, S58), and one manager (S78) provided only endline survey data.

Full administration data is available for all settings (attendance at EYCP training sessions, hub meetings, and support visits). All seven settings were invited to participate in interviews or focus groups, and four took part (two from each cohort). N09 and N17 completed separate manager and Champion interviews; S17 completed a focus group with the Champion and a practitioner (buddy) and a separate interview with the manager. S17 took part in a Champion and manager focus group.

All setting Champions were observed at least once as they attended hub meetings, and baseline and endline ITERS-3 data was collected from all settings at baseline and endline. Table 42 describes the data collected from the good practice settings.

Table 42: Data collected from the good practice settings

Setting ID	Survey data	Admin data	Interview/focus group	Training observation	ITERS-3 audit/observation data
N06	Incomplete Champion: Baseline only Manager: Did not complete	Yes	No	Yes	Yes
N09	Yes	Yes	Yes: Champion and manager interviews	Yes	Yes
N17	Incomplete Champion: Baseline only Manager: Baseline only	Yes	Yes: Champion and manager interviews	Yes	Yes
S17	Yes	Yes	Yes: Champion and practitioner focus group and manager interview	Yes	Yes
S54	Incomplete Champion: Baseline only Manager: Did not complete	Yes	No	Yes	Yes
S58	Incomplete Champion completed baseline and endline Manager: Baseline only	Yes	No	Yes	Yes
S78	Incomplete Champion: Baseline only Manager: Endline only	Yes	Yes: Manager and Champion focus group	Yes	Yes

Identifying good/best practice setting features

Utilising all data collected listed above in the IPE study, seven themes had emerged in identifying the setting features for good/best practice.

Theme 1: Does setting type and size matter?

Slightly over half of the total nurseries (57/101) are private or voluntary settings; therefore, maintained settings are slightly overrepresented in this sample. This suggests that maintained nurseries were more likely to be identified by mentors as demonstrating good/best practice. Type of setting may impact access to resources, training, and specialist support which may influence the skills and knowledge of Champions and the wider team. For example, maintained settings with a specialist language provision within the setting enables them to receive direct training from Speech and Language Therapists:

> *'A lot of it [training] is around SEN and about speech because of the DSP, so obviously we all we all have that training as well.'* (MS17)

Maintained setting N09 are a research school and Stronger Practice Hub which may provide greater opportunities to engage with high-quality research. However, private and voluntary settings were also nominated suggesting that setting type does not necessarily determine success in engaging with and implementing the EYCP training.

The seven nominated settings had an average of 21 two-year-olds at baseline ranging from 8 to 47 which is consistent with the total sample. There are likely pros and cons associated with smaller and larger settings. For example, settings with a greater number of children likely have more staff which may facilitate finding cover for Champions to attend training.

> *'Because we are based in school, we can cover in like internally. So because there's other staff that can come down. It's not ideal when we have to cover, because they're not specially trained staff. But we haven't had to pay for anybody else to come in. So we're already here. So we've just moved like one to one to come and support, for example. So no, it's not cost anything.'* (MN09)

However, the manager from setting S17 which has a separate provision for two- and three-year-olds with only eight two-year-olds enrolled believes a smaller number of staff may have facilitated implementation, as it provides more opportunity for flexibility in delivery and a more consistent approach.

> *'When you've got seven staff in a nursery, it gets a lot more difficult because obviously you've got more children, more level of need, more probably things on the timetable.'* (MS17)

> *'Originally I had a buddy, but we are small team so all staff [have] been doing it.'* (CS78)

Theme 2: What are their training attendance records like?

Managers were expected to attend one full day and one half day of face-to-face training, and one online training session. Champions were expected to attend one full day and one half day of face-to-face training sessions, six online training modules, nine hub meetings, and five support visits. Attendance among managers varied with four attending all three required sessions and three attending just one. Champions' attendance (with the exception of N06 who missed four online modules) was generally very good. Most attended at least five out of six online training sessions, at least eight out of nine hub meetings, and all Champions attended both face-to-face training and all five support visits. Maintained settings generally had higher attendance particularly for the manager. Three out of four managers who attended the in-person training were from maintained settings.

Theme 3: What is their setting environment like measured by ITERS at baseline and endline? What are the changes in environments between baseline and endline?

All settings had higher total and subscale ITERS-3 scores at endline compared to baseline suggesting all settings made improvements to their language environments following training. S17 in particular saw large increases in their Language and books scores and N17 and S78 in their Interaction scores. The manager at N09 discussed making changes to the language environment by implementing interactive reading practices: *'It's also made for a richer environment for them day to day, because of the interactive storytelling they love'* (MN09).

The manager at S78 also discussed making changes to routines to include more interaction: *'In snack time, we talk about the food and give the children more choices—encourage them to self-serve or have conversations around the food and use that time to have more conversations with children'* (MS78).

As shown in Table 43, baseline ITERS-3 scores varied across settings, indicating differences in the quality of provision prior to the intervention. Overall scores ranged from 3.0 to 5.42 (all settings mean score = 3.76), suggesting that settings were operating at broadly minimal to good quality levels at baseline. Variation was also evident across subscales. For the Language and books subscale, scores ranged from 2.8 to 5.3 (all settings mean score = 3.74), while Interaction scores ranged more widely from 1 to 6 (all settings mean score = 3.75), indicating substantial differences in the quality of staff-child interactions across settings. Some settings (for example, N09) demonstrated relatively high baseline scores across domains, suggesting more established practices prior to the intervention. In contrast, others (for example, N17 and S17) had lower baseline scores, particularly in Interaction, indicating greater scope for improvement. This variation in baseline scores suggests differing starting points across settings, which is important to consider when interpreting subsequent changes and may partly explain differences in the magnitude of improvement observed at endline.

Table 43: ITERS at baseline and endline for good practice settings

Setting ID	Baseline ITERS-3 Language and books subscale	Baseline ITERS-3 Interaction subscale	Baseline ITERS-3 overall score	Endline ITERS-3 Language and books subscale	Endline ITERS-3 Interaction subscale	Endline ITERS-3 overall score
N17	3.3	1	3	4.6	4	4.57
N09	5.3	6	5.42	6.66	7	6.7
N06	3.4	4	3.5	4.2	4	4.1
S58	4.16	4	4.14	6.1	6	6.1
S17	2.8	4	3.4	6.3	6	6.2

S78	4.17	3	4	6.5	7	6.6
S54	4.6	3	4.4	6.2	6	6.14
All settings (*N* = 102)	3.74	3.75	3.76	5.46	5.65	5.48

Theme 4: What went well within those settings? (e.g. staff engagement, managers role and organizational support)

Interviews with Cohort 1 best practice settings (N09 and N17) revealed they were able to have regular meetings and ad hoc discussions about the programme which facilitated implementation of the EYCP and use of ShREC strategies in the settings. The managers at settings N17 and N09 discussed the willingness and openness of staff to embrace change as key to success.

> *'I found it very successful; the staff was really on board, they tried very hard, they well supported each other, they were open to thinking about changes and adapting and try to find ways to make things happen, I think it's very successful.'* (N17)

> *'There isn't really anyone that would kind of put their foot down and say, I'm not doing that just because it's new. They all are open to learning and growing as practitioners. So it's been really successful because everyone's wanted it to be and have been really interested in it and actually believed in it.'* (N09)

All interviewed settings discussed cascading the EYCP training and ShREC approach to the wider team. For some, this was relatively formal including creation of training materials adapted from EYCP resources from the Padlet. Others discussed a more ad hoc or naturalistic approach suggesting aspects of the training such as structured reflection were embedded into daily practice: *'It's just something that we had the ShREC approach sheets out, and they would just sort of do it throughout the day'* (CN17).

One manager commented on the importance of providing Champions time and space to both attend training and to discuss and reflect on what they had learned.

> '*I just kind of felt that my role was there to support the Champion to make sure that you know she had the time to attend the sessions, that she had the space to then reflect on it, and we would reflect on what she'd done. Therefore, we reflected on that together, and kind of she could help make sense of it.*' (MN09)

Settings also discussed cascading training to parents which was an unanticipated consequence of the programme.

> *'We've also been able to do parent workshops around the ShREC approach, etc. And so I'm hoping it's also empowered the parents to feel more capable to kind of identify any potential language difficulties and to be able to support the child to develop the language better.'* (MN09)

> *'Some children with emerging SEN, [the Champion] has done some training sessions with parents so they can try at home the interactive conversation. Because we have seen a big difference in the children so if the parents can do it at home, that's a really good joint approach.'* (MN78)

Theme 5: What challenges have they encountered? What were the solutions?

Good practice settings encountered two main challenges: difficulties around releasing staff to attend training and supporting colleagues to engage with the ShRec strategies, and changing ingrained patterns of behaviour and established approaches to language learning and interactions with children.

The Champion in setting N17 commented that she herself found it challenging to change her own behaviour but that it became easier over time. The Champion at N09 struggled at first to get across the importance of using ShREC to colleagues who hadn't attended training.

> *'When you're part of it in person. It has a bigger impact than someone else just telling you about it. So for me, trying to make sure that the rest of my team are as passionate as I am and are following through with the strategies, has been one of my challenges, my personal challenges.'* (CN09)

She highlighted the support and advice of her mentor as being particularly helpful in resolving this.

> *'M3 has been really supportive and given me some tips on how to manage that and having a buddy system. So it's not so much pressure on me as well. And she said, even if you manage to just get one other person who's really implementing it like you, then that's a win in itself.'* (CN09)

The manager at N09 discussed low staff ratios as a potential barrier, although she did not consider this to be a major barrier at their setting, and noted that they mitigated this challenge by having good management structures and practices in place.

> *'They [staff] have very clear routines... So really kind of emphasising that importance of teamwork and kind of being dynamic and vigilant within the nursery, so that you can mitigate those times when it's all going a bit bonkers.'* (MN09)

For setting S17, this was evidently more of an issue:

> *'It's a nightmare because ratios; we work on a minimum ratio basis. So if [the Champion's] gone, she has to be replaced. There's no surplus of staff here. So, yeah, so it was. It was any more. I think it would have been more difficult.'* (MS17)

At S17, both the Champion and manager also talked about difficulties releasing staff which was made more challenging due to their distance from other settings in the hub. This made attending face-to-face hub meetings very difficult. This was resolved when the mentor created hybrid hub meetings allowing the Champion to attend online. Although both the Champion and manager acknowledged that they would prefer her to attend in person if it was possible.

Theme 6: If any, what are the observed outcomes within these settings (for example, child level, staff level, setting level, parents level)?

Settings N17, N09, and S78 spoke in general terms about the positive impact the EYCP was having on the children's language development:

> *'I do feel like in the last six months a lot of the language and the progression of children's speech has come on a lot since really incorporating the EYCP into our daily routine.'* (CN17)

> '*Anecdotally, they can see just through being with children every day how their conversational skills are developing and improving.*' (MN09)

At setting S78, the Champion talked specifically about the impact on children with low levels of language, and that staff are focusing more on interacting with those children. *'I think that we are spending a lot more time with children who are not communicating. We have seen a difference in their confidence as they can see that we are sharing their attention and valuing in what they are saying'* (CS78).

The Champion at N09 also observed positive impacts on a child learning English as an Additional Language. The Champion at N17 believes that relationships between staff and children have improved following increased focus on shared attention and relational pedagogy.

> *'I think it's also made relationships between some of the children and adults much better as well, because, especially with the sharing of the attention. It's just you're present, you don't necessarily have to say anything but the fact that you're there. They appreciate that.'* (CN17)

At setting N09, the manager remarked that the training made staff feel valued and instilled a feeling of positivity and enthusiasm for their work; the Champion at the setting spoke about how the EYCP has given her more confidence and helped her to be more reflective in her practice and approach to language learning. In setting S78, the Champion also discusses in general terms changes to practice and approach and the impact on the wider team.

> *'We can see that the staff are being more mindful about what they are saying, definitely giving the children more time to respond. People are being more mindful and intentional with their teaching.'* (CS78)

Theme 7: Continuity *of the programme beyond the training (if any)*

Setting S17 have plans to extend the implementation of the EYCP into the three-year-old provision by having the Champion train staff who work with that cohort.

> *'…because we've got 2-year-olds and a three-year-old provision. So what we want to do is bring more of it into the three-year-old provision, but we just haven't had the time to do the training but in September. So CS17 is going to train all the three-year-old staff.'* (MS17)

Setting S78 are focusing on ensuring the training is included in inductions for new staff and delivering refresher training during the period when lots of new children are joining the setting: *'it is something that might be included in our induction process so staff understand it from the beginning'* (CS78).

> *'This time of year we have lots of children joining and leaving so we have had a little staff meeting to refresh what we are doing during the settling in period. The ShREC approach is something we will encourage staff to do.'* (MS78)

Ideal types of best/good practice settings

Analysis of the nominated good/best practice settings suggests that success in implementing the EYCP and ShREC approach is not determined solely by structural characteristics such as setting type or size. Rather, effective implementation appears to result from an interaction between structural enablers, leadership practices, professional culture, and implementation processes.

Structural factors: Necessary but not sufficient

Maintained settings were somewhat overrepresented among nominated settings and often benefitted from greater access to specialist expertise, research networks, and internal staffing capacity. These features appeared to facilitate high training attendance and opportunities for structured reflection. However, private and voluntary settings were also successfully nominated, suggesting that structural advantage alone does not determine implementation quality.

Similarly, setting size presented both advantages and constraints. Larger settings may have greater flexibility to release staff for training, while smaller settings may benefit from relational coherence and ease of cascading practice across a small team. The data suggests that neither size nor type alone predicts success; instead, how these structural features are mobilised is more important.

Leadership and implementation infrastructure

Across settings, strong managerial support for the Champion role was a defining feature of good practice. Where managers actively protected time for training, reflection, and implementation, Champions reported greater confidence and sustained behaviour change. High attendance at training, particularly among Champions, was also characteristic of strong implementers.

Regular opportunities for reflection—whether formal (scheduled meetings) or informal (ongoing discussions)—facilitated embedding of ShREC strategies into everyday practice. Mentoring support and adaptive solutions (for example, hybrid hub meetings) mitigated contextual challenges such as travel or staffing pressures.

Professional culture and collective buy-in

The most consistent feature across successful settings was a culture of openness to change. Staff willingness to adapt established patterns of interaction and engage reflectively with their practice emerged as central to effective implementation. In these settings, ShREC strategies became embedded into daily routines rather than treated as an add-on initiative.

Cascading approaches varied (formal presentations versus organic modelling and discussion), but in all cases, good practice settings ensured that the wider team was meaningfully engaged. Some extended this to parents, strengthening consistency across home and setting environments.

Observed outcomes

All nominated settings demonstrated improvements in ITERS-3 scores from baseline to endline, suggesting measurable enhancements in language environments such as increased availability and use of books and literacy materials; more frequent adult-child verbal interactions; greater use of descriptive language, questioning, and modelling vocabulary; more responsive and sustained interactions; improved turn-taking and conversational exchanges; and increased sensitivity to children's cues and communication attempts. Other reported impacts included:

- increased child conversational confidence;
- improved engagement among children with low language levels and EAL learners;
- stronger staff-child relationships;
- greater staff intentionality and reflective practice; and
- enhanced professional confidence and enthusiasm.

Sustainability

Good practice settings demonstrated forward planning for sustainability, including embedding ShREC into induction processes, extending training to other age groups, and refreshing practice during transition periods.

Overall, the analysis suggests that good/best practice settings can be understood as those that successfully align structural capacity, leadership support, professional culture, and reflective implementation processes.

Cost

Based on information shared by the delivery team in the cost workshops and using the EEF costing guidance (EEF, 2023), we estimated the cost of delivering the EYCP programme in year 1 of EYCP delivery as it was in the project. This included calculating the costs for three models:

- Cohort 1 delivery to 22 settings using a paired mentor model with each mentor pair supporting four to seven settings (mean 5.5 settings).
- Cohort 2 delivery to 49 settings with each mentor supporting five to eight settings (mean 6.1 settings).
- The expected ongoing delivery model for 56 settings with eight mentors each supporting seven settings as planned originally.

To deliver the project in a setting for a second and third year, there were expected to be only minimal additional costs. The delivery team felt that ongoing training and mentor support would not be necessary for delivery beyond the first year, and that implementing EYCP strategies would become part of normal practice rather than need additional resourcing or training. The offer of optional termly alumni peer support online meetings as well as ongoing access to the Padlet resources were the only expected support needs for settings. We assumed one day each for the delivery team project leads would be required to facilitate the peer support meetings and included this in the costs for Year 2 and Year 3 (an additional £700 for each year). Access to Padlet resources was not expected to require additional funding while the project continued offering any training or support to schools.

Costs for the programme delivery are set out in Table 44. Settings in this trial did not pay anything towards the programme and were provided with additional funds to cover staff to attend training and implement the programme, which are included in the programme cost calculations. The only prerequisite resources settings required were access to a digital camera for filming and sharing nursery practice and a computer or laptop with camera and microphone to participate in online training and meetings and to access the online resources.

Table 44: Costs of EYCP programme delivery for Cohort 1 and 2 as delivered and for an ongoing delivery model

			Cohort 1 (2023/2024) 22 settings	**Cohort 2 (2024/2025) 49 settings**	**Expected ongoing delivery— 56 settings**
Mentor Model			*Working in pairs, each pair supporting average 5.5 nurseries (£5,950 per mentor)*	*Working individually, each mentor supporting average 6.1 nurseries (£7,875 per mentor)*	*Working individually, each mentor supporting 7 nurseries (£7,875 per mentor)*
	Type of cost	Item	£	£	£
YEAR 1 DELIVERY	**Delivery team personnel costs for programme delivery**				

	Set-up cost: staffing	Delivery team personnel costs for delivery of programme including: project management, administration, training delivery, and ongoing support to nurseries	£12,550	£12,550	£12,550
	Set-up cost: staffing	8 mentors providing training and support to nurseries. Time is to run hub meetings, conduct ITERS and mentoring visits, provide ongoing support, and attend training. Cohort 1 based on 17 days per mentor. Cohort 2 and ongoing model based on 22.5 days per mentor.	£47,600 (£5,950 x 8)	£63,000 (£7,875 x 8)	£63,000 (£7,875 x 8)
	Delivery team travel costs for delivery				
	Set-up cost: travel	Travel to 2 x in-person training events for 2 x delivery team	£80	£80	£80
	Set-up cost: travel	Travel for eight mentors to attend training, hub meetings, and visits. Cohort 1: 14 days cap x £20 Cohort 2 and ongoing model: 15 days cap x £20	£2,240 (14 x £20 x 8)	£2,400 (15 x £20 x 8)	£2,400 (15 x £20 x 8)
	Delivery team other costs				
	Set-up cost: resources	Provision of book, two posters, and folder per setting at £10.25 per setting	£225.50	£502.25	£574
	Set-up cost: venue	Venue hire and catering for training events	£3,557	£8,424	£8,424
	Ongoing cost: licences	Zoom licence (£471), Padlet licence (£96), and Vimeo licence (£108) for hosting online resources and delivery of training	£675	£675	£675
	Setting costs				
	Set-up cost: cover to attend training	Cover costs* for EYCP Champion to attend training, engage with	£18,067.50	£40,241.25	£45,990

			mentor, and plan for EYCP delivery (assuming 29 hours training and 1 hour x 26-week planning, preparation, assessment time is 7.5 days and £109.50 per day backfill cost)				
		Set-up cost: cover to attend training	Cover costs* for senior leader to attend 1.4 days of training (£109.50 per day backfill cost)		£3,372.60	£7,511.70	£8,584.80
		Set-up cost: travel costs to attend training	Travel costs for EYCP Champion to attend 7 in-person events and for senior leader to attend 2 in-person events (£20 per event)		£3,960	£8,820	£10,080
			Total cost Year 1 overall		£92,328	£144,204	£152,358
		Total cost Year 1 per setting	£4,197	£2,943	£2,721		
		Total cost Year 1 per child	£198	£139	£128		
YEAR 2 DELIVERY		**Delivery team personnel for delivery**					
		Ongoing cost:	1 day planning and delivery time for two delivery team project leads to host 3 x 1-hour online alumni events		£700	£700	£700
			Total cost Year 2 overall		£700	£700	£700
		Total cost Year 2 per setting	£32	£14	£13		
		Total cost Year 2 per child	£1.50	£0.67	£0.59		
YEAR 3 DELIVERY		**Delivery team personnel for delivery**					
		Ongoing cost:	1 day planning and delivery time for two delivery team project leads to host 3 x 1-hour online alumni events		£700	£700	£700
			Total cost Year 3 overall		£700	£700	£700
		Total cost Year 3 per setting	£32	£14	£13		
		Total cost Year 3 per child	£1.50	£0.67	£0.59		

	TOTAL COSTS			
	Total cost for delivery over 3 years	£93,728	£145,604	£153,758
	Cost per setting for 3 years' delivery	£4,260.35	£2,971.51	£2,745.68
	Cost per setting per year of delivery (averaged across 3 years)	£1,420.12	£990.50	£915.23
	Cost per child per year of delivery—assuming mean number of 2-year-old children per setting across both cohorts (21.22)**	£66.92	£46.68	£43.13
	Cost per child per year of delivery—assuming maximum number of 2-year-old children in a participating setting (60)	£23.67	£16.51	£15.25
	Cost per child per year of delivery—assuming minimum number of 2-year-old children in a participating setting in (4)	£355.03	£247.63	£228.81

* Cover costs here are based on the rate of £109.50 per day which was the rate provided by the delivery team to settings for backfill costs to enable practitioners to attend training and implement the programme. This was offered to all settings, and most settings invoiced for this. Typical costs for EY supply (level 3) are £80–£120 per day.
**Mean calculated using expected number of 2–3-year-old children in settings as reported in Memorandum of Understanding when signing up to project.

For the Cohort 1 delivery model, the cost per setting for delivery over three years was £4,260.35 (£1,420.12 per year). For the Cohort 2 delivery model, the cost per setting for delivery over three years was £2,971.51 (£990.50 per year). For the proposed ongoing delivery model, assuming seven settings per mentor, the three-year cost per setting would be slightly reduced at £2,745.68 (£915.23 per year). The main cost difference between Cohort 1 and Cohort 2 is due to the delivery model with individual mentors supporting more settings each in Cohort 2.

Over three years of delivery, three cohorts of children would be expected to receive the EYCP programme. To estimate the number of children that would benefit from the programme per year, we used the mean number of two-year-old children expected to attend participating study settings as provided at the start of the project (mean = 21.22 children, range 4 to 60 children). We therefore assume that 64 children per setting would benefit from the programme over three years. For the Cohort 1 delivery model, the cost per child would be £66.92. For the Cohort 2 delivery model, the cost per child would be £46.68. For the ongoing delivery model, the cost per child would be £43.13. However, as the cost of the programme delivery would be the same for all nurseries, smaller settings would have a much higher rate per child compared to larger settings. For example, for the ongoing delivery model, for the trial setting with the smallest group of two-year-olds (4), the cost per child per year would be £228.81, while for the setting with the largest group of two-year-olds (60), the cost per child would be £15.25.

The time commitment for practitioners to deliver the programme over the first year (as delivered) and expected time commitment for the subsequent two years is outlined in Table 45. In the first year, the time included for practitioner training and mentor support is based on discussion of the project requirements with the delivery team. Estimates for the ongoing planning time practitioners need to deliver EYCP have been taken from the EYCP Champion survey responses and have been informed by discussions with practitioners during interviews. For the second- and third-year delivery, it is assumed that the EYCP Champion would continue delivery of the programme as well as cascading programme content wider in the setting. EYCP Champions created an action plan for how they would continue delivery into Year 2 at their final hub meeting, often including cascading and further training and support. We have assumed that additional cascading of the programme would be continued as an hour per week across 26 weeks, similar to the planning time in Year 1 of delivery. As the programme provided professional development and strategies to use throughout usual nursery practice, delivery of the activities in the classroom did not require any additional time or resourcing over usual practice.

Table 45: Total time required of nursery personnel for training, preparation, and delivery

Type of cost	Staff member	Activity	Activity time (hours)	Number of staff	Overall time (hours)
YEAR 1					
Start-up	EYCP Champion	EYCP training (1.4 days* in person + 6 hours online)	17	1	17
Start-up	EYCP Champion	Attend 9 x 1 hour hub meetings (5 in person with 1 hour travel included and 4 online)	14	1	14
Start-up	EYCP Champion	Engage with mentor support visits (5 x 1 hour)	5	1	5
Start-up	Senior Leader/Manager	EYCP training (1.4 days) including travel	11	1	11
Ongoing	EYCP Champion	Weekly preparation for delivery and cascading of EYCP strategies including meetings with line managers and supporting buddy Reported in survey as average of 1.5 hours per week**	39	1	39
Ongoing	EYCP Champion	Using EYCP strategies in classroom—part of normal practice (no additional time)	0	0	0
YEAR 2 and YEAR 3—time requirements per year					
Ongoing	EYCP Champion	Providing training and support to other staff in implementation of EYCP strategies (assuming 1 hour per week over 26 weeks)	26	1	26
Ongoing	EYCP Champion and other practitioners	Using EYCP strategies in classroom—part of normal practice (no additional time)	0	0	0
Ongoing	EYCP Champion	Attend optional online EYCP alumni peer network sessions (assume 3 x 1 hour per year)	3	1	3

Total amount of time	Start-up	EYCP Champion	36 hours		
	Ongoing (each year)	EYCP Champion	39 hours		
	Start-up	Senior Leader/Manager	11 hours		

*1 day assumed to be 7.3 hours.
** Specific weekly hours of preparation time provided by 22/87 unique responses of EYCP Champions, excluding an outlier reporting 30 hours of prep time (average time spent 1.5 hours, median 1 hour, and range 0 to 4 hours). Actual number likely to be lower with those not needing prep time not completing the question.

Data was collected about any additional costs that settings incurred to deliver the EYCP programme through the EYCP Champion survey and interviews. In the survey, 21 respondents indicated not having additional costs, while 12 respondents indicated that they had made additional purchases. Of those, the majority had purchased additional books for the setting (*n* = 11), while three settings had purchased some furniture (for example, a sofa for the book corner, mats for reading), and four settings had purchased other resources including cushions and pictures for their book corner or resources to go with books (for example, puppets, word cards). In interviews, practitioners felt that these additional costs were to improve the nursery provision as a result of the EYCP training, rather than being necessary to deliver the programme, and therefore these costs have not been included in the costs above.

Conclusion

Table 46: Key conclusions

Key conclusions
1. Children in EYCP intervention nurseries made no additional months' progress in receptive and expressive vocabulary, on average, compared to children in the waitlist-control nurseries. This result has a low to moderate security rating.
2. Among children living in the areas of highest deprivation, those in EYCP intervention nurseries made two fewer months progress in receptive vocabulary, on average, compared to those in waitlist-control nurseries. These results may have lower security than the overall findings because of the smaller number of children.
3. Children in nurseries that implemented EYCP with high fidelity and who attended the setting regularly appeared to make more progress in receptive vocabulary compared to children in control nurseries. This result is less secure than the main finding due to the smaller sample size.
4. The EYCP programme improved practitioners' language-related pedagogical knowledge. This finding has low security because of the small number of practitioners.
5. Practitioners reported increased confidence in supporting children's language and communication, along with changes in classroom practice through implementation of the EYCP strategies using the ShREC approach and interactive book reading. The programme was perceived to benefit all children, regardless of their language background.

Impact evaluation and IPE integration

Evidence to support the logic model

The logic model for the Early Years Conversation Project (EYCP) set out a clear causal chain: high-quality professional development (PD) delivered with fidelity would lead to improved practitioner knowledge, confidence, and practice, which in turn would enhance the language environment in nurseries and ultimately improve young children's language and communication outcomes. The evaluation provides strong evidence for most aspects of this model, although the final link of improved child language outcomes within the trial timeframe was only partially supported. The paragraphs below further substantiate the positive impacts of programme delivery.

Fidelity of professional development and programme delivery

There is strong evidence from the IPE that the PD training was delivered as intended and was well received and valued by nursery practitioners. Attendance by EYCP Champions and by setting managers/leaders at the training was high across the required elements, and observations confirmed that training was delivered in line with the pre-specified manualised structure, with mentors providing consistent, high-quality support throughout hub meetings and in-setting visits. The resources, multi-strand learning design, and opportunities for video-based reflection functioned as specified in the logic model, supporting both engagement and understanding across settings.

The EYCP Champion role was well defined and generally delivered with fidelity, aligning closely with the logic model. Champions attended the full range of training, engaged actively in mentoring, and implemented ShREC strategies in their daily interactions with children. Most also cascaded training to other staff, either formally through working with a 'buddy', team meetings or informally through modelling practice. EYCP Champions felt that the programme structure and their role and responsibilities as EYCP Champion were clear and appropriate. These findings provide support for the role of the Champion as a central mechanism for embedding practice change.

In contrast, the setting manager/EY lead role was less clearly defined in practice particularly regarding providing support for the Champion and EYCP implementation which was more variably implemented. While many managers provided strong facilitative support enabling attendance at training and mentoring meetings, allocating time for reflective discussions, and reinforcing expectations, others played a more limited or flexible role. Interviews with EYCP Champions highlighted the importance of leadership support with implementation especially in regard to being able to cascade to other staff and for

the EYCP Champion to feel supported and confident to make changes to practice. The expectations and responsibilities for setting managers/EYFS leads may require further clarification and specification to ensure more consistent and facilitative support is provided.

Practitioner knowledge, confidence, and nursery practice

The evaluation provides consistent evidence supporting the practitioner-focused mechanisms of the logic model. Practitioners in intervention settings reported increased confidence in supporting children's language development, and scores on the Observing Language Pedagogy (OLP) tool indicated higher levels of pedagogical knowledge for the intervention group. Champions described tangible changes in their classroom practice, including greater use of the ShREC strategies, more intentional pedagogy around language skills, and more frequent multi-turn conversations and frequent use of interactive book reading. These findings were triangulated across surveys, interviews, observations, and administrative data, offering good evidence that the PD successfully produced the intended behavioural and pedagogical changes.

Environmental audits (ITERS-3) of the nursery before and after implementing the EYCP also showed improvements consistent with enhanced language environments. Together, these findings indicate that the intermediate outcomes of the theory of change were achieved.

Child outcomes

The ultimate goal of the programme to improve young children's receptive and expressive vocabulary was not supported by the trial's primary impact findings. Across all children, the EYCP produced no additional months of progress in receptive or expressive vocabulary, or in more general language and communication skills compared to the waitlist-control group. Subgroup analyses similarly showed no evidence of impact for children from disadvantaged areas, those with English as an additional language, or children with lower or higher baseline language scores. However, compliance analyses offer partial support for the causal mechanism: settings that engaged more fully with the programme (at both setting and child levels) demonstrated stronger effects, with optimal compliers showing gains equivalent to around two months' additional receptive vocabulary progress. This indicates that the logic model pathway from improved practitioner practice to improved child outcomes may operate under high fidelity conditions but was not consistently realised across all settings during the trial. It may be that the timeframe of the programme delivery is not long enough for the practice changes at the practitioner level to feed through to changes in children's language skills except in those settings most dedicated to the programme. The intervention timeframe may have been insufficient for effects on child outcomes to fully emerge, and a longitudinal follow-up could provide additional insight into whether impacts become more apparent over time. Variation in children's attendance is also likely to have resulted in substantial differences in intervention dosage. For example, a child attending the setting for 15 hours per week would require approximately 16 months to receive the same level of exposure as a child attending 30 hours per week over an eight-month period. Furthermore, exposure to the intervention is likely to have been influenced not only by attendance but also by the extent to which EYCP practices were cascaded throughout the setting. Although cascading was not positioned as a core component of the EYCP programme in the logic model, it was widely adopted in practice. However, because cascading was neither tightly structured nor consistently supported within programme delivery, there was considerable variation in how it was enacted across settings. Given that only around half of the settings were cascading the programme beyond one additional staff member (aside from the EYCP Champion), children's exposure to the intervention may have been more limited than would otherwise be expected.

In the qualitative data, practitioners reported children being more engaged in classroom conversation, more confident in having conversations, and showing more enjoyment in the conversation activities as a result of the programme. Practitioners reported that children exhibited increased confidence in communication, developed stronger relationships with practitioners, and demonstrated positive behavioural changes. These were considered additional, unanticipated benefits of the programme. Triangulation of findings from the CACE analysis and the IPE indicates that setting-level implementation fidelity was generally high. This suggests that the effects observed in the CACE analysis may have been driven primarily by variation in child-level exposure to the intervention, rather than differences in implementation across settings. Taken together, these findings suggest that dosage, or exposure to the intervention, is likely to be an important mediator, both in

relation to the timeframe required for effects to emerge and in explaining the positive effect size observed in the CACE analysis.

Revisions to the logic model

Based on the study data, the logic model can be considered broadly accurate in its assumptions regarding the inputs, outputs, and proximal outcomes (PD delivery, practitioner knowledge and confidence, and improvements to practice and environment). However, the pathway from practitioner change to child language outcomes appears weaker than expected, at least over an 8-month implementation period and for children aged two years. This suggests that the logic model may need revision to reflect:

- a longer timescale for child-level effects to emerge;
- greater specification of the setting manager/EYFS lead role, which appears to facilitate implementation but was enacted inconsistently;
- the necessity of high dosage/exposure of children to the intervention, and high setting compliance for effects to materialise, highlighting a potential threshold mechanism not explicitly accounted for in the original logic model; and
- the need to consider measurement challenges with two-year-olds, where assessment reliability and developmental variability may dampen detectable effects within short windows.

Interpretation

The finding that EYCP showed no overall impact on children's receptive or expressive vocabulary is consistent with a broader body of early years intervention research. Large-scale evaluations of early language programmes frequently report small or null average effects, particularly when implemented under routine conditions rather than tightly controlled efficacy trials. For example, early years evaluation trials funded by the Education Endowment Foundation have repeatedly shown that early language and communication programmes tend to yield modest effects on average, often in the range of 0–3 months' additional progress. The Education Endowment Foundation Early Years Toolkit (Higgins *et al.*, 2017) similarly reports that communication and language approaches in early years (although children are 3 years and older) typically have small positive effects, with considerable variability depending on implementation quality. The uncertainty interval in the present trial (ranging from approximately -3 months to +2 months on the primary outcome) reflects a common challenge in early childhood intervention research: relatively small true effects combined with substantial variability in children's development and measurement noise. For the current trial, assessing 2-year-olds is particularly complex due to rapid developmental change, variability in attention, and contextual influences on performance, all of which can attenuate detectable effects.

Although the primary outcome measure, BPVS-III, is widely used, robust, and standardised, it captures a relatively narrow domain of language. If EYCP influenced broader communicative behaviours (for example, interaction quality, conversational turn-taking, narrative skills), these changes may not have translated into detectable gains on a single-word receptive vocabulary test within the study timeframe. Similarly, secondary outcome measures such as the CDI-III and WellComm measure expressive vocabulary and comprehensive language skills but may still be insufficiently sensitive to detect incremental improvements in interactional or pragmatic language use. Previous research highlights that vocabulary growth is strongly influenced by the quality and quantity of adult-child interaction (Hoff, 2013; Law *et al.*, 2017b; Pan *et al.*, 2005; Snow *et al.*, 1998). However, measurable vocabulary gains may require sustained, high-intensity exposure over time. Short- to medium-term interventions may therefore struggle to produce large detectable changes on standardised tests, especially among very young children.

In terms of benefits to subgroup children, the absence of measurable benefits for children from disadvantaged areas (IDACI 1–3), those with EAL, or those at the lowest baseline language levels warrants careful consideration. Whereas evidence consistently shows that socio-economic disadvantage is associated with gaps in early language development (Hart and

Risley, 1995; Law *et al.*, 2017b), interventions targeting disadvantaged children often show smaller-than-expected effects when delivered universally (Higgins *et al.*, 2017). Similarly, for children with EAL, improvements in English receptive vocabulary may not fully capture broader language competence or development in the home language. Research suggests that bilingual development follows distinct trajectories, and gains may not immediately appear on monolingual standardised assessments (Hoff, 2021). It is noteworthy that findings from the compliance analyses indicate that increased adherence to the EYCP model at setting level with higher child-level programme exposure was associated with stronger intervention effects. In particular, children who were more exposed to the intervention in settings classified as optimal compliers made progress equivalent to approximately two additional months in receptive vocabulary.

While the EYCP did not demonstrate an overall statistically significant impact on children's vocabulary outcomes, practitioners in EYCP intervention nurseries consistently scored higher on the language-related pedagogical knowledge scale (OLP) compared to practitioners in waitlist-control nurseries. This suggests that the programme may have had a positive influence on practitioner knowledge and understanding of language development and pedagogy. Previous literature indicates that high-quality training can improve practitioner knowledge, which in turn may influence classroom practice and child outcomes (Sylva *et al.*, 2004). However, these findings must be interpreted cautiously due to the small number of practitioners included in the analysis, the results did not reach statistical significance, and the security of the findings is therefore considered low to moderate.

The implementation and process evaluation provides important contextual insight into the efficacy findings. In the IPE study, practitioners reported increased confidence in supporting children's language and communication development following participation in EYCP. In this case, reported increases in confidence align with the observed (though not statistically significant) improvements in language-related pedagogical knowledge among intervention practitioners. Together, these findings suggest that the programme was successful in influencing practitioner-level outcomes, even if this did not translate into consistent, measurable child-level gains across all settings.

Limitations and lessons learned

Limitations of the impact evaluation

Missing data and attrition

Attrition at child level was 21.6% overall, with higher attrition in the intervention arm (25.3%) than the control arm (17.8%). Children were excluded from analysis where baseline data was missing (21 cases), which also reduced statistical power. At baseline, this data was often missing due to the child's inability to understand and complete any of the assessment. Although missingness was assumed to be missing at random and unlikely to have biased the results in favour of one group, it reduced the effective sample size and may have weakened the detectable signal of impact.

Measurement challenges with 2-year-olds

The trial relied on standardised assessments of children's language ability at aged two years. At this age, there is large variability in children's language development, self-regulation, and social skills which adds to the difficulty in collecting valid and accurate research data. At the first assessment stage, children were often brand new to the nursery setting. The use of external researchers who were unfamiliar to the child led to a lack of child engagement in the assessments for some children despite the support of nursery staff. The nature of the standardised assessment, and particularly for the BPVS primary outcome, which needed to continue until a ceiling level was reached, meant that the assessment could be long and the child sometimes lost concentration or was unwilling to continue with the assessment. Assessment scores did not always represent the highest score the child could have achieved. Some nursery practitioners commented that they did not feel this type of assessment was the most appropriate to use with children of this age group. The trial did also collect parents' reports of expressive vocabulary to counter some of these issues, but completion rates at endline were lower than for the BPVS, and the long list of words from the measure may have put some parents off completing it fully limiting its ability as a reliable measure.

Limitations of split cohort design

Use of the split cohort design maximised the potential sample size through delivery of the trial over two years in two separate cohorts. However, due to this design, the delivery of the programme was not identical across both cohorts. In the first cohort, the mentor support model meant that mentors worked paired with a second mentor in supporting settings, while in the second cohort mentors mostly worked individually, each supporting their own settings. The programme content also changed as a result of practitioner feedback and the observation of the delivery team during the first cohort, with changes to the order of the training sessions and a greater emphasis on interactive book reading as a way to use the ShREC approach. The political context also changed during the delivery of the project with a universal offering of free childcare for two-year-olds being brought in for the second cohort, changing the delivery context and potential sampling for the second cohort. Analysis of both cohort data separately as a sensitivity analysis indicated that results were similar across the two cohorts and that the changes made between the cohorts are unlikely to have influenced the conclusions.

Limitations of the process and implementation evaluation

Much of the process evaluation data (confidence, cascading activities, perceptions of impact) relied on practitioner self-report. This potentially introduced bias through participants exhibiting social desirability bias or through inaccurate judgement of the practitioner's own ability which may have influenced the accuracy of the data collected this way.

Although settings were selected at random for many of the practitioner interviews, settings were able to choose whether to take part or not. This may have resulted in some selection bias with the sample potentially being skewed towards nurseries who viewed the programme more positively and who were therefore more willing to take part. The addition of settings purposively chosen as best practice settings may again have added to this. Acknowledgement should be made that the IPE does not represent all settings who took part in the EYCP.

For the ITERS data collection, which was collected by mentors with a close working relationship with the practitioner by the end of the project, their knowledge of the practitioner and their relationship could have led to scoring their observations in a more positive light. However, this may have been minimised by nursery practitioners not being given their ITERS scores at either timepoint.

Lessons learned

CACE analyses showed that optimal compliers demonstrated child-level gains of approximately two months' receptive vocabulary progress—indicating that high dosage, high-fidelity implementation can produce measurable impact. Future implementations of the programme should therefore emphasise structures that enable and incentivise compliance.

Further consideration of how to reliably assess 2-year-olds' language ability at scale should be made due to the challenges with administering the BPVS-III and WellComm measures with two-year-olds in the trial and the low response rate from parent report measures at trial endline. Future trials may benefit from: alternative or additional observational measures, assessments completed by a nursery practitioner familiar to the child and/or increased assessor familiarisation time in the nursery before assessment.

Longer trial timeframes may be needed to detect child outcomes. Practitioner-level impact was evident within the eight-month delivery period; however, child outcomes may require a longer period to manifest—particularly given the age group, the variability in exposure due to children's attendance patterns, and the incremental nature of language development. Extending follow-up into the subsequent year could provide a clearer test of the causal pathway. It may also be that the next cohort of children in the nursery would be those to get the full benefit of the practitioner's change in practice. As the programme training and reflection spanned the eight-month period of the trial, there was minimal time for practitioners to implement changes in practice from later in the programme, and children would only have had limited exposure to these changes. It may be that greater practitioner change in practice using knowledge from the full programme is not evident until the school year after the trial and collecting data from children in the following cohort of two-year-olds could change this more fully.

Generalisability of the results

The sample included 102 nurseries across London, spanning PVI, maintained, and school-based settings, with a large and diverse population of 2-year-olds. This broad base supports the generalisability of findings to similar early years contexts particularly those operating in urban, mixed provision systems.

However, generalisability is constrained by:

- the variability of how setting managers/early years leads delivered their supporting role in the project;
- moderate levels of attrition and missing data across different data collection aspects of the programme;
- the programme's reliance on high compliance for impact in children's outcomes; and
- contextual differences between London settings and smaller, rural, or less diverse areas.

Accordingly, the results should be interpreted as most applicable to settings with similar demographic characteristics and resource conditions. The qualitative findings on practitioner development and implementation facilitators/barriers are likely to generalise more widely than the child-level impact findings.

Future research and publications

It would be valuable for further research to explore the following questions that it wasn't possible to explore in this study:

- How does EYCP practice change become embedded and sustained over time in nurseries beyond the original programme period?
- Do the EYCP impacts on children emerge in subsequent years (for example, at age 3 or school entry) or for the next cohort of 2-year-old children in the setting?
- Are there any particular features of nursery settings (for example, staffing stability, leadership engagement) which moderate the strength of the causal chain or which facilitate the high compliance with programme activity that seem necessary to impact on children's outcomes?

The evaluation team plan to write academic papers in peer-reviewed journals in collaboration with the delivery team which will be based on the results of this research.

References

Altman, D. G. and Bland, J. M. (2005) Treatment allocation by minimisation. Bmj, 330.7495.843.

Ang, L., Grenier, J., Harmey, S., James, F., Prendergast, M. (2025) 'Supporting early years professional development to enhance high-quality conversational interactions in early childhood care and education settings: the ShREC approach'. *Early Years*, 10.1080/09575146.2025.2498717.

Braun, V. and Clarke, V. (2006) 'Using thematic analysis in psychology'. *Qualitative Research in Psychology*, 3(2), pp. 77–101. DOI: 10.1191/1478088706qp063oa.

Burchinal, M. R., Pace, A., Alper, R., Hirsh-Pasek, K., Golinkoff, R. M. (2016) Early language outshines other predictors of academic and social trajectories in elementary school. Presented at Assoc. Child. Fam. Conf. (ACF), Washington, DC, July 11–13.

Dale, P. S., Kachergis, G., Weber, A. M., and Frank, M. C. (2023) The CDI-III. The MacArthur-Bates Communicative Development Inventories: User's guide and technical manual, 3rd Edn. Baltimore, MD: Brookes Publishing Co.

Department for Education (2021) Education recovery support Guidance. Available at: **https://www.gov.uk/government/publications/education-recovery-support.** (accessed 19 May 2026)

Department for Education (DfE) (2023a) Statutory framework for the early years foundation stage. Available at: https://www.gov.uk/government/publications/early-years-foundation-stage-framework—2. London: Department for Education. (accessed 13/10/2026)

Department for Education (DfE) (2023b) Early years stronger practice hubs guidance. Available at: https://www.gov.uk/government/publications/early-years-stronger-practice-hubs/early-years-stronger-practice-hubs. London: Department for Education. (accessed 13 October 2026)

Dimova, S., Ilie, S., Brown, E.R., Broeks, M., Culora, A. & Sutherland, A., 2020. The Nuffield early language intervention. Education Endowment Foundation.

Dimova, S., Scott, M., Scandone, B., Sciarra, A., Rezaian, M., and Bury, J. (2022) Hanen Learning Language and Loving It (Hanen LLLI) Evaluation protocol. London: Education Endowment Foundation.

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. Available at: https://discovery.ucl.ac.uk/id/eprint/10170459/1/Talking%20Time%20Nuffield%20Final%20Report%20May%20%202023%20for%20public%20dissemination%20.pdf (accessed 13 October 2026)

Dunn, L. M. and Dunn, L. M. (1981) PPVT-R 'Peabody picture vocabulary test—revised' Manual. *Circle Pines, MN: American Guidance Service*.

Dunn, L. M., Dunn, D. M., and National Foundation for Educational Research (NFER) (2009) The British Picture Vocabulary Scale—Third edition. London: GL Assessment.

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

Education Endowment Foundation (EEF) (2019) Classification of the security of findings from EEF evaluations.

Education Endowment Foundation (EEF) (2022a) Statistical Analysis Guidance for EEF Evaluations. London: EEF.

Education Endowment Foundation (EEF) (2022b) Implementation and process evaluation guidance for EEF evaluations. London: EEF.

Education Endowment Foundation (EEF) (2023) Cost evaluation guidance for EEF Evaluations. London: EEF.

Education Endowment Foundation (n.d.) Early Years measures database. Review of measures for children aged 0–6 in language, literacy, numeracy, and social and emotional skills. London: EEF. Available at: https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluation/eef-outcome-measures-and-databases/early-years-measures-database-2/early-years-measures-database (accessed 13 October 2026)

Education Endowment Foundation (n.d.) Early Years Toolkit: Communication and language approaches. London: EEF. Available at: https://educationendowmentfoundation.org.uk/education-evidence/early-years-toolkit/communication-and-language-approaches. (accessed 13 October 2026)

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., and Reznick, J. S. (2007) MacArthur-Bates communicative development inventories. Baltimore, MD: Brookes Publishing.

GOV.UK (2023a) Data set 2: Children registered and providers by provider type. Education provision: children under 5 years of age, Reporting year 2023—Explore education statistics. Available at: https://explore-education-statistics.service.gov.uk/data-tables/fast-track/4891da30-91d7-44b5-75ef-08db786e904b. (accessed 31 July 2023) London: HM Government.

GOV.UK (2023b) Upcoming changes to childcare support. Available at: https://www.childcarechoices.gov.uk/upcoming-changes-to-childcare-support/#:~:text=The%20upcoming%20expansion&text=This%20means%20that%3A,to%203%2Dyear%2Dolds. (accessed 31 July 2023) London: HM Government.

Graham, J. W., Olchowski, A. E., and Gilreath, T. D. (2007) 'How many imputations are really needed? Some practical clarifications of multiple imputation theory'. *Prevention Science*, *8*(3), 206-213.

Harms, T., Cryer, D., Clifford, R. M., and Yazejian, N. (2017) 'Infant/Toddler Environment Rating Scale, third edition'. *New York, NY: Teachers College Press*.

Hart, B. and Risley, T. (1995) Meaningful differences in the everyday experience of young American children. Paul H Brookes Publishing.

Haux, T., Butt, S., Rezaian, M., Garwood, E., Woodbridge, H., Bhatti, S., Woods Rogan, R., and Paul, G. (2022) The early years workforce: recruitment, retention, and business planning. Department for Education. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1072062/SCEYP_thematic_report-_April_2022.pdf. (accessed 24 May 2023)

Higgins, S.; Katsipataki, M.; Villanueva-Aguilera, A.B.; Coleman, R.; Henderson, P.; Major, L.E.; Coe, R.; Mason, D.(2017) Sutton Trust—Education Endowment Foundation Early Years Toolkit | Communication and language approaches, London: EEF. https://educationendowmentfoundation.org.uk/early-years/toolkit/communication-and-language-approaches/ (accessed 13 October 2026)

Hoff, E. (2013) 'Interpreting the early language trajectories of children from low-SES and language minority homes: implications for closing achievement gaps'. *Dev. Psychol.* 49, 4–14. DOI: 10.1037/a0027238.

Hoff, E. (2021) 'Why bilingual development is not easy'. *Advances in child development and behavior*, *61*, 129–167. https://doi.org/10.1016/bs.acdb.2021.03.002

Husain, F., Morris, S., Basi, T., and Bristow, T. (2020) Manor Park Talks. NatCen: Independent Evaluator. Available at: https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/manor-park-talks(open in a new window) (accessed 10 February 2026)

Jin, M., Polis, A., and Hartzel, J. (2020) _Minirand: Minimization Randomization_. R package version 0.1.3, <https://CRAN.R-project.org/package=Minirand>.

Law, J., Charlton, J., and Asmussen, K. (2017a) Language as a child wellbeing indicator. London: The Early Intervention Foundation.

Law, J., Charlton, J., Dockrell, J., Gascoigne, M., McKean, C., and Theakston, A. (2017b)Early Language Development: Needs, provision, and intervention for preschool children from socio-economically disadvantaged backgrounds. EEF Report. London: EEF. (accessed 13 October 2026)

Law, J., Charlton, J., McKean, C., Watson, R., Roulstone, S., Holme, C., Gilroy, V., Wilson, P., and Rush, R. (2020) Identifying and Supporting Children's Early Language Needs. Newcastle, UK: Newcastle University.

Law, J., Charlton, J., Wilson, P., Rush, R., Gilroy, V., and McKean, C. (2023) 'The development and productivity of a measure for identifying low language abilities in children aged 24–36 months'. *BMC pediatrics*, *23*(1), 495. https://doi.org/10.1186/s12887-023-04079-x

Lehrl, S., Evangelou, M., and Sammons, P. (2020) The home learning environment and its role in shaping children's educational development. Sch. Eff. Sch. Improv. 31, 1–6. DOI: 10.1080/09243453.2020.1693487.

Lindorff, A., Sylva, K., Ereky-Stevens, K., and Joseph, A. (2022) Coaching Early Conversation, Interaction and Language (CECIL) Impact evaluation. Sutton Trust. Available at: https://www.suttontrust.com/wp-content/uploads/2022/03/CECIL-Evaluation-Oxford-14.02.2022.pdf. (accessed 13 October 2026)

Markussen-Brown, J., Juhl, C. B., Piasta, S. B., Bleses, D., Højen, A., and Justice, L. M. (2017) 'The effects of language- and literacy-focused professional development on early educators and children: A best-evidence meta-analysis'. *Early Childhood Research Quarterly*, 38, pp. 97–115.

Mathers, S. (2020) Observing Language Pedagogy (OLP): Developing and piloting a contextualised video-based measure of early childhood teachers' pedagogical language knowledge. Doctoral thesis (Ed. D.), UCL (University College London).

Mathers, S. J. (2021) 'Using video to assess preschool teachers' pedagogical knowledge: explicit and higher-order knowledge predicts quality'. *Early Childhood Research Quarterly*, 55, 64–78. DOI: 10.1016/j.ecresq.2020.10.010.

Mathers, S., Malmberg, L. E., Lindorff, A. M., and Gardiner, J. (2022) Early educators' knowledge of early language pedagogy: How can it be measured and does it matter for child language outcomes? Oxford: University of Oxford.

Melhuish, E. C., Phan, M. B., Sylva, K., Sammons, P., Siraj-Blatchford, I., and Taggart, B. (2008) 'Effects of the Home Learning Environment and Preschool Center Experience upon Literacy and Numeracy Development in Early Primary School'. *Journal of Social Issues*, 64: 95-114. DOI: 10.1111/j.1540-4560.2008.00550.

Mol, S. E. and Neuman, S. B. (2014) 'Sharing information books with kindergartners: the role of parents' extra-textual talk and socioeconomic status'. *Early Childhood Res. Q.* 29, 399–410. DOI: 10.1016/j.ecresq.2014.04.001.

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*(2017), 285–308.

Pan, B. A., Rowe, M. L., Singer, J. D., Snow, C. E. (2005) 'Maternal Correlates of Growth in Toddler Vocabulary Production in Low-Income Families', *Child Development*, Volume 76, Issue 4, July, Pages 763–782, **https://doi.org/10.1111/1467-8624.00498-i1**

Public Health England (2020a) Best start in speech, language and communication: Guidance to support local commissioners and service leads. Public Health England. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/931310/BSSLC_Guidance.pdf. (accessed 13 October 2026)

Public Health England (2020b) Early Language Identification Measure and Intervention: Guidance handbook. London: Public Health England, Department of Health and Social Care, Department for Education.

Puma, M., Bell, S., Cook, R., Heid, C., and Lopez, M. (2005) Head Start impact study: First year findings. Administration for Children and Families.

Qi, X., Menzies, V., Siddiqui, N., Ismail, N., & Oughton, R. (2023). Independent evaluation of the Early Years Conversation Project (EYCP): A two-armed cluster randomised waitlist controlled trial. Evaluation Protocol. Education Endowment Foundation. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/EYCP-protocol_final.pdf?v=1791904645. (accessed 13 October 2026)

Robinson-Smith, L., Whiteside, K., Togerson, C., Fairhurst, C., Podmore, D., Davill, T., Bell, K., Qi, X., Leggett, H., Elliott, L., Hewitt, C., Baird, K., Menzies, V., Torgerson, D., and Ainsworth, H. (2023) Independent evaluation of Maths Champions in nursery to develop children's early numeracy: A two-armed cluster randomised controlled trial. Education Endowment Foundation.

Roy, P., Kersley, H., and Law, J. (2004) The Sure Start Language Measure Standardisation Study: Report to the Sure Start Unit London. Available from Newcastle University c/o School of Education, Communication and Language Sciences, University of Newcastle, Newcastle-upon-Tyne.

Schwab, J. F. and Lew-Williams, C. (2016) 'Language learning, socioeconomic status, and child-directed speech'. *Wiley Interdiscip Rev Cogn Sci*. 7(4):264-275. DOI:10.1002/wcs.1393.

Sénéchal, M., Ouellette, G., and Rodney, D. (2006) The misunderstood giant: On the predictive role of vocabulary to reading. In S. B. Neuman and D. Dickinson (Eds.) *Handbook of Early Literacy* (Vol.2; pp.173–182). New York: Guilford Press.

Snow, C. E., Burns, M. S., and Griffin, P. (1998) Preventing reading difficulties in young children. Washington, DC: National Academy Press.

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., and Taggart, B. (2004) The effective provision of pre-school education (EPPE) project technical paper 12: The final report-effective pre-school education. IOE: London.

Sylva, K., Joseph, A., Ereky-Stevens, K., and Lindorff, A. (2021) Responding to covid requirements in an early childhood language intervention RCT. European Early Childhood Education Research Association online conference.

Tracey, L., Torgerson, C., Welch, C., Fairhurst, C. M., Gridley, N., Dysart, E., and Qi, X. (2019) Efficacy of the ParentChild+ programme: trial evaluation protocol. Education Endowment Foundation.

Tracey, L., Torgerson, C., Dysart, E., Fairhurst, C., Gridley, N., and Welch, C. (2022) Efficacy Trial of the ParentChild+ programme: Evaluation Report. Education Endowment Foundation.

WellComm (2010) A Speech and Language Toolkit for Screening and Intervention in the Early Years. GL Assessment.

Wooles, N., Swann, J., and Hoskison, E. (2018) Speech and language delay in children: a case to learn from. *British Journal of General Practice*, *68*(666), 47–48. https://doi.org/10.3399/bjgp17X694373

Appendix A: EEF cost rating

Figure 1: Cost Rating

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

Appendix B: Security classification of trial findings

Rating	Criteria for rating			Initial score		Adjust		Final score
	Design	**MDES**	**Attrition**					
5	Randomised design	<= 0.2 (=0.19)	0-10% (=9%)					
4	Design for comparison that considers some type of selection on unobservable characteristics (for example RDD, Diff-in-Diffs, Matched Diff-in-Diffs)	0.21 - 0.29	11-20%					
3	Design for comparison that considers selection on all relevant observable confounders (for example Matching or Regression Analysis with variables descriptive of the selection mechanism)	0.30 - 0.39	21-30%	3		Adjustment for threats to internal validity		
2	Design for comparison that considers selection only on some relevant confounders	0.40 - 0.49	31-40%			**[-1]**		2
1	Design for comparison that does not consider selection on any relevant confounders	0.50 - 0.59	41-50%					
0	No comparator	>=0.6	>50%					

Threats to validity	Threat to internal validity?	Comments
Threat 1: Confounding	Moderate	The evaluation was well designed with balanced intervention and control groups (conducted by an independent member of the team). The baseline ELIM score had an imbalance of -0.09 which implies moderate risk. This was controlled for in the regression model. Sensitivity analyses were conducted using both baseline scores on primary outcome results displaying no significant discrepancies.
Threat 2: Concurrent Interventions	Low	Practitioners in both groups mentioned participating in language-focused CPD training during baseline including Early Talk Boost, Early Years Professional Development Programme (EYPDP), Makaton etc. 31 respondents in the intervention group stated participation in other CPD programmes compared to 22 in the control group. However, participation in other EEF trials was an inclusion criteria and take up of these seemed low prior to participation in the trial.
Threat 3: Experimental effects	Low	The waitlist-control design minimises the risk of the control group behaving differently. There is no further evidence available in the report to assess this.
Threat 4: Implementation fidelity	Low	Implementation fidelity and compliance were well defined and aligned with the logic model. According to the IPE section, implementation fidelity was high among participants. Minimal compliance was observed for 262 children across 52 nurseries and optimal compliance was observed for 49 children across 28 nurseries (total of 346 children).

Threat 5: Missing Data	Moderate	Missing data is moderate (21.6%) with some difference between the treatment groups (25.3% in intervention and 17.8% in control). The analyses accounting for missing data through multiple imputation are similar to the complete case analyses.
Threat 6: Measurement of Outcomes	Low	The measurement tests are valid and reliable for this age group. The evaluator provides information on administration related issues, but these could not have differentially impacted children in the treatment groups. The tests were externally marked by individuals blinded to group allocation.
Threat 7: Selective reporting	Low	The trial was registered and the analysis aligns with the protocol and SAP. Deviations in terms of the baseline measure and primary outcome measure have been explained and justified in the report. Additional analyses with the initially proposed measures are also included.

- **Initial padlock score**: [3] Padlocks – This was a well conducted cluster randomised trial with MDES at randomisation of 0.25 and 21.6% attrition

- **Reason for adjustment for threats to validity:** [- 1] Padlocks – There is evidence consistent with an imbalance in baseline assessment and differential attrition between the intervention and control groups which suggests the impact estimate could be overestimated.

- **Final padlock score**: initial score adjusted for threats to validity = [2] Padlocks

Appendix C: Effect size estimation

Table 1: Effect size estimation

Outcome	Unadjusted difference in means (95% CI)	Adjusted difference in means (95% CI)	Intervention N (missing)	Intervention Variance	Control N (missing)	Control Variance	Pooled Variance
Endline BPVS-III	+0.65 (−1.72, +3.02)	−0.50 (−3.14, +2.16)	346 (85)	252.103	361 (71)	264.028	258.13
Endline CDI-III score	+3.36 (−0.72, +7.44)	+0.71 (−2.12, +3.54)	314 (117)	700.788	306 (126)	640.147	670.82
WellComm	−0.13 (−0.49, +0.23)	−0.23 (−0.68, +0.24)	350 (103)	5.9997	361 (82)	5.8508	5.92
Naming (OLP)	+0.83 (−0.24, +1.90)	+1.20 (−0.06, +2.46)	34 (17)	6.5383	17 (33)	1.5294	5.13
Interpretation (OLP)	+2.53 (+0.36, +4.70)	+3.39 (+0.57, +6.22)	34 (17)	35.3868	17 (33)	1.8603	24.91
Perceiving strategies (OLP)	+9.50 (−2.62, +21.62)	+10.00 (−2.58, +22.57)	34 (17)	504.9713	17 (33)	355.6202	460.12
BPVS-III: EAL	+0.39 (−3.51, +4.29)	−0.70 (−4.33, +3.33)	90 (27)	157.9107	95 (18)	204.5207	181.7
BPVS-II: Bottom 20%	−0.95 (−5.04, +3.14)	−3.07 (−6.53, −1.12)	49 (33)	137.4566	71 (23)	101.0427	115.75
BPVS-III: Top 20%	−3.99 (−9.63, +1.65)	−2.81 (−8.44, +2.71)	68 (12)	256.0009	67 (4)	292.2411	273.94
BPVS-III: IDACI 1–3	−1.30 (−5.15, +2.55)	−1.70 (−5.85, +2.41)	108 (39)	222.0217	156 (35)	276.3026	255.28

Further appendices

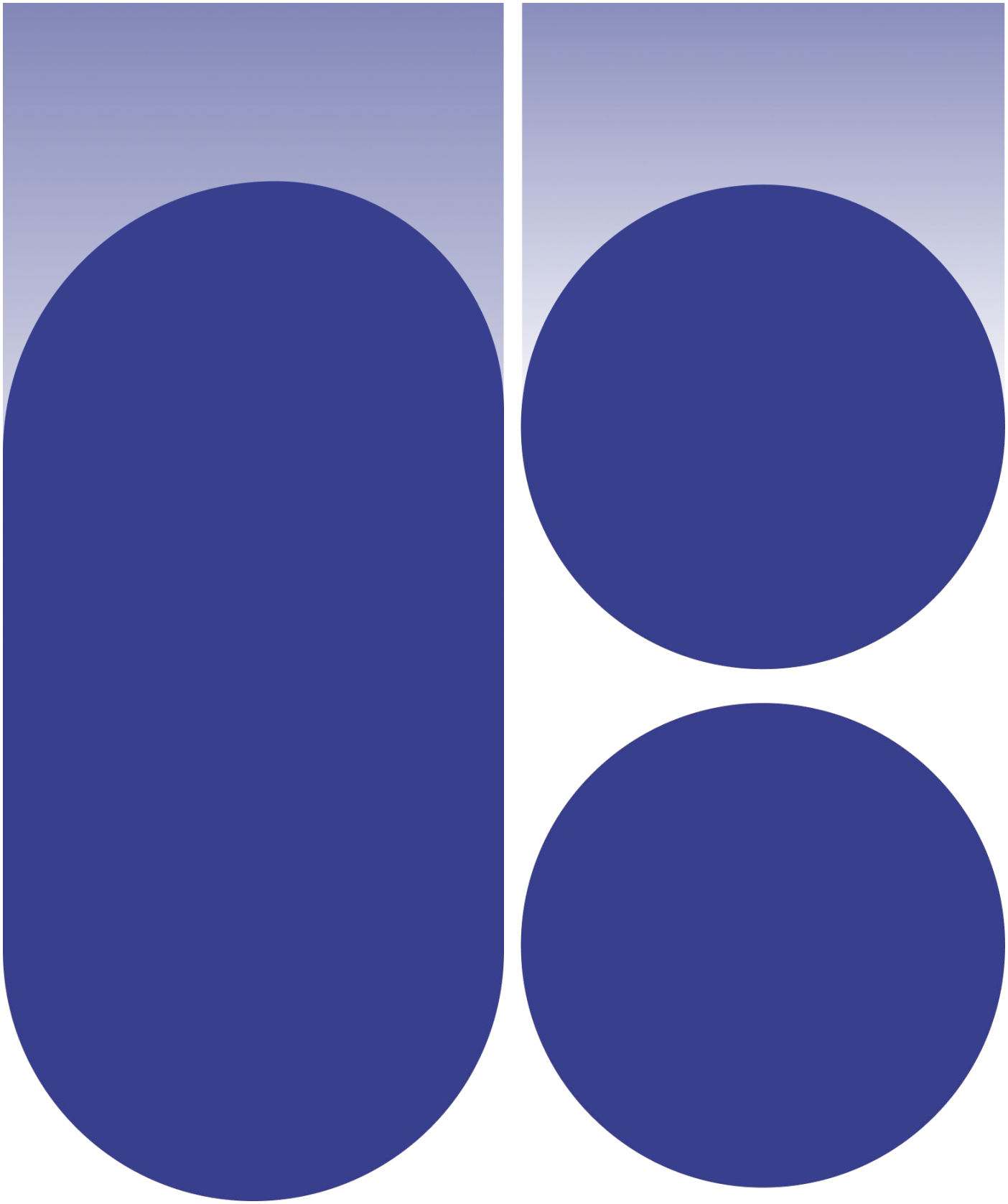
Appendix D – F can be found published in an accompanying document '*Further Appendices'*.

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