Coaching Early Conversation, Interaction and Language (CECIL) Impact evaluation

Ariel Lindorff, Kathy Sylva, Katharina Ereky-Stevens and Allen Joseph

31 January 2022

Acknowledgments

The broader Coaching Early Conversations, Interaction and Language pilot project was managed by the Sutton Trust and funded by Esmée Fairbairn Foundation, the Lindsell Foundation and the Sutton Trust. The broader CECIL pilot project encompasses the interventions conducted in Hackney and Nottinghamshire, an implementation and process evaluation conducted by the Institute for Employment Studies, and the impact evaluation conducted by our research team at the University of Oxford.

Our Oxford research team would like to thank the Speech and Language Therapy teams in Nottinghamshire and Hackney, as well as Anneka Dawson and Clare Huxley from the Institute for Employment Studies team working on the parallel implementation and process evaluation, for the friendly, collegial and collaborative partnership working across organisations that made this project possible, and of course to the Sutton Trust and in particular Laura Barbour for facilitating and overseeing that partnership working throughout.

We would also like to offer special thanks to Philip Dale for permission to use the anglicised version of the CDI-III, and to James Law for support in the early stages of this work and for permission to use the ELIM word list before publication.

Finally, we would also like to formally acknowledge the contribution of Dr Katharina Ereky-Stevens, who acted as Principal Investigator during the period of baseline data collection and setting allocation while Dr Ariel Lindorff was on parental leave. Dr Ereky-Stevens also led on the development of the instrument for coding video clips of practice.

Table of Contents

Executive summary

Background

There has long been compelling evidence that greater socioeconomic disadvantage is associated with weaker language skills at school entry (Duncan, Magnusun & Votruba-Drzal, 2017; Melhuish & Gardiner, 2018; Sylva et al. 2004). Further, gaps in language skills increase throughout primary school (Dearden et al., 2011). This underscores the importance of interventions that target children before school entry in order to close gaps associated with socioeconomic disadvantage.

Although manualised interventions have proved popular in many local authorities (LAs) and Early Years settings, there are other LAs and settings that seek more tailored support for practitioners who wish to enhance children's language development. This impact evaluation investigated the effects on child and practitioner outcomes of coaching support to practitioners delivered by Speech and Language Therapists. This small-scale impact evaluation was part of a broader project aiming to assess the feasibility of scaling up a programme of coaching-centred intervention and corresponding evaluation.

The interventions whose evaluation we report here took place in Nottinghamshire and Hackney, in early years settings in the Private, Voluntary and Independent (PVI) sector. Interventions targeted children who were 2-3 years old at the start of the 2020-21 school year. Interventions in the two LAs were planned and delivered by the LA Speech and Language Therapy teams who worked directly with the practitioners in each intervention setting. As a result of the pandemic, implementation of the interventions involved a mix of face-to-face and in-person support to participating settings/practitioners.

Alongside the impact evaluation led by our Oxford research team, the Institute for Employment studies conducted a parallel implementation and process evaluation (Dawson, Huxley and Garner, 2022).

Research questions

Our impact evaluation focused on the following questions:

1. What is the effect of each intervention on child speech and language?
2. What is the effect of each intervention on practitioner outcomes (observed practice, confidence and professional knowledge)?
    a. How do practitioner outcomes relate to child outcomes?

Additional feasibility questions included:

3. What early language child assessments are appropriate as pre- and post-test measures for a future larger-scale study?
4. What measure(s) of observed practice, practitioner confidence and professional knowledge are appropriate as pre- and post-test measures for a future larger-scale study?

Methods

20 settings in Nottinghamshire and 20 settings in Hackney were allocated to Early Starter (intervention) and Late Starter (control, with a version of each intervention delivered after the evaluation) conditions in each LA using minimisation to achieve balance between groups of settings

on characteristics such as deprivation which are known to influence practice and consequently children's language development.

Children's language was measured via parent-report. Parents were asked to tick the words their children used on a list of 50 words in Autumn 2020, and a list of 100 words in Summer 2021, either in an online form or on paper.

Practitioner confidence and skills were measured via a self-report questionnaire developed by the research team, with input from the Institute for Employment Studies research team who also had access to practitioner survey data to inform the process evaluation. The same items asking about practitioners' confidence in supporting children's language development and their skills (in terms of interactions with children to support language development) were included in a baseline survey in Autumn 2020 and a post-intervention survey in Summer 2021.

There was severe attrition in the number of parent responses over the course of the evaluation, likely in part due to the pandemic. In Autumn 2020, there were 178 valid parent responses in Nottinghamshire and 117 in Hackney. In Summer 2021, at least in part due to challenges associated with the pandemic (e.g. setting closures, staff absence, heavier workloads for practitioners in the context of which distributing links/materials to parents constituted an additional burden on setting managers), only 45 parents provided responses in Nottinghamshire and 22 in Hackney.

There was also considerable attrition in practitioner responses: 41 responded in Autumn 2020 and only 19 in Summer 2021 in Nottinghamshire, while 54 responded in Autumn 2020 and 20 in Summer 2021 in Hackney.

Because of the small sample size as a result of this attrition, results must be interpreted with caution, and more sophisticated approaches to analysis that would have been appropriate in a larger sample were simply not possible. We relied on a combination of simple inferential statistics (t-tests, regression) and descriptive statistics to inform our findings as a consequence of this limitation.

Results

Child language

After taking into account children's language (number of words) in Autumn 2020, in Nottinghamshire children in the Early Starter group scored about 11 points higher on average than those in the Late Starter group in Summer 2021. This difference was borderline statistically significant, suggesting that there may be an effect of the intervention despite challenges caused by the pandemic both to the intervention itself and the evaluation.

In Hackney, no significant difference was found between the Early and Late Starter groups in terms of children's language development by Summer 2021. This does not mean that the intervention had no effect, only that the evaluation was not able to detect an effect, which may be a result of the very small analytical sample size.

Practitioners' confidence and skills

No significant differences were found between the Early and Late Starter groups in terms of change in confidence and skills based on practitioner self-report between Autumn 2020 and Summer 2021. This does not mean that the interventions in Nottinghamshire and Hackney had no effect, but the

evaluation was not able to detect effects given a very small sample of practitioners (15 in Nottinghamshire and 14 in Hackney) for whom data could be matched across the two time points.

Observations of practice

Originally, the intention was to observe practice during research visits to settings. This was not possible due to the pandemic. Instead, we used video clips shared by practitioners and developed a coding instrument to observe practice via these video recordings, with reasonable evidence of inter-rater agreement, i.e., close or exact agreement on most items. Unfortunately, few settings submitted post-intervention videos, and it was not possible based on the sample obtained to make comparisons from Autumn 2020 to Summer 2021 nor across Early and Late Starter groups.

Recommendations regarding feasibility and scale-up of evaluation methods

While it is clear that there were circumstantial hurdles to the implementation of this evaluation, we did learn some valuable lessons to inform future, larger-scale evaluations of similar interventions with an emphasis on coaching.

- There is plenty of previous evidence of the validity of parent-report to measure child language (e.g. Dale, 1996; Feldman et al., 2005; Law et al., 2020), and despite the various limitations of the present study, we also found some evidence of the validity of parent-reported child vocabulary as well as the potential of such measures to be sensitive enough to detect an effect (Sylva et al 2021). We would recommend the use of parent report *alongside the researcher or practitioner assessment of child language, including dimensions of language beyond vocabulary,* that were not feasible to include in the present study. The combination of such measures with parent-reported child vocabulary would provide a more complete understanding of child language development and any differences in child language due to intervention.
- There is previous evidence of the validity of practitioner self-report to measure pedagogical skills and knowledge (e.g. Mathers, 2021), even though in this study there were issues with the response rate/sample size and no effect of the interventions in Hackney or Nottinghamshire were detected. We would recommend the use of practitioner self-report in a larger-scale study, but if possible this should be complemented with an alternative measure based on researcher observation (even if in a subsample rather than the entire sample in order to reduce costs) for the sake of triangulation.
- If using video-recordings of practice, we would recommend a structured approach that ensures the comparability and utility of video clips via a tightly defined set of instructions and the gathering of explicit information about why practitioners chose the activities and children that they include in videos. We would also recommend careful consideration of the strategies used to transfer videos, as this proved to be a substantial technological impediment to submitting videos for some settings (e.g. in some cases videos would not upload to the secure sharing platforms due to limitations in the settings of internet connections/computers, despite the best efforts of practitioners and the research team).

Background

There has long been compelling evidence that greater socioeconomic disadvantage is associated with weaker language skills at school entry (Duncan, Magnusun & Votruba-Drzal, 2017; Melhuish & Gardiner, 2018; Sylva et al. 2004). Further, gaps in language skills increase throughout primary school (Dearden et al., 2011). This underscores the importance of interventions that target children before school entry in order to close gaps associated with socioeconomic disadvantage. There have been several successful interventions targeting children 3-5 years old (Dockrell et al., 2016; Fricke et al., 2013), but many of these have consisted of structured interventions based on detailed manuals for practice and specific materials for children. The Education Endowment Foundation (EEF) has evaluated several preschool language interventions and found that only a handful led to positive results on child outcomes when subjected to rigorous RCT evaluations (e.g. Sibieta et al., 2016). There are more promising effects of continuing professional development intervention on practice outcomes, for example the EEF trial on the Using Research Tools to Improve Language in the Early Years (URLEY) programme (Wright, Carr, Wiese, Stokes, Runge, Dorsett, Heal, Anders, 2020) found higher quality scores in intervention settings compared to control. Finally, there have been small-scale studies such as that of McDonald and colleagues (2015) who found improvement in language-supporting practices after a brief training course led by Speech and Language Therapists. Although manualised interventions have proved popular in some local authorities (LAs) and schools, there are other LAs and settings that seek more tailored approaches to support them in enhancing children's language development. This impact evaluation investigated the effects on child and practitioner outcomes of individual coaching support to Early Years practitioners carried out by Speech and Language Therapists. This small-scale impact study was part of a broader project aiming to assess the feasibility of scaling up a programme of coaching-centred intervention and implementing a larger scale evaluation.

All settings participating in this research were in the Private, Voluntary and Independent (PVI) sector. Interventions targeted children who were 2-3 years old at the start of the 2020-21 school year. Interventions in the two participating local authorities (LA), Hackney and Nottinghamshire, were planned and delivered by the LA Speech and Language Therapy teams who worked directly with the practitioners in each intervention setting.

The intervention in Nottinghamshire, 'Let's Interact', was a local adaptation of the Hanen Program for Early Childhood Educators 'Learning Language and Loving It' training for practitioners (Weitzman & Greenberg, 2002). The intervention aimed to enhance the skills of practitioners in supporting language development, for example by encouraging turn-taking in conversations. The programme of intervention included a combination of group training sessions, individual coaching sessions based on videos made by participating practitioners and feedback from Speech and Language Therapists on these videos, discussions via telephone or video chat, language- lead network meetings and project network sessions.

The intervention in Hackney, Launchpad for Language (L4L; Children's Integrated Speech and Language Therapy Service for Hackney and The City, n.d.), was intended to be a universal approach to supporting all children to reach their communication potential based on principles from the Early Years Foundation Stage guidance from the Department for Education (DfE). Settings participating in the L4L intervention were offered a menu of packages intended to be tailored to their strengths and

needs, with packages including child interventions, staff interventions and parent workshops as three key strands of support.

The evaluation of these interventions was conducted in partnership with the Institute for Employment Studies, which led on the process evaluation strand of the research, while our research team at the University of Oxford led the impact evaluation strand. Both interventions are described in greater detail in the process evaluation report by Dawson, Huxley and Garner (2022).

Research questions

This impact evaluation aimed to answer the following primary research questions:

1. What is the effect of each intervention on child speech and language?
2. What is the effect of each intervention on practitioner outcomes (observed practice, confidence and professional knowledge)?
    a. How do practitioner outcomes relate to child outcomes?

An important purpose of the impact evaluation was also to inform potential future larger-scale evaluations of similar types of interventions. With this in mind, additional feasibility questions included:

3. What early language child assessments are appropriate as pre- and post-test measures for a future larger-scale study?
4. What measure(s) of observed practice, practitioner confidence and professional knowledge are appropriate as pre- and post-test measures for a future larger-scale study?

Impact evaluation approach

In order to address the above research questions, our overall approach to evaluation prioritised the specific nature of interventions devised and implemented by Speech and Language Therapists. We drew on existing instruments but tailored the versions we used to the priorities and contexts of the interventions themselves. In order to accomplish this, our adaptations of instruments were informed by conversations via email and video-chat with the Speech and Language Therapy team in each LA as well as materials shared by them to clarify their priorities and foci in delivering the interventions and assessing practitioners' progress.

Methods

Recruitment

At the beginning of the study, recruitment involved a joint effort between the Speech and Language Therapists delivering the intervention in each LA, the IES, and the University of Oxford research team. Speech and Language Therapists made the initial approach to settings, as they were familiar with the context and leading on the intervention element which involved the greatest time commitment from – and provided the greatest immediate potential benefit for – settings. Setting managers willing to participate in the intervention and evaluation signed an initial expression of interest. Speech and Language Therapy teams selected the settings that would be able to participate, after which the IES and Oxford teams sent information and consent forms specific to the implementation and process evaluation (IES) and the impact evaluation (Oxford). For the Oxford team, this involved sending versions of impact evaluation information and consent tailored for setting managers, practitioners and parents/carers. Setting managers distributed the Oxford information and consent forms for practitioners (intended to be 2 per setting) and parents/carers

(intended to be for 10 children per setting) via their usual communication channels, e.g. WhatsApp groups, email or paper distribution. The Oxford team provided all settings with both paper and online versions of the relevant information and consent so that participants could choose their preferred way to respond. We included a video created by each LA Speech and Language Therapy team to introduce the intervention and evaluation in the online parent information, as well as a reader-friendly flyer with information about the intervention and evaluation for parents who preferred paper copies. All interested settings were promised an intervention, but the experimental group were scheduled to receive the intervention in the year 2020-21 ('Early Starters') and the control group ('Late Starters') in the following year.

Inclusion criteria for children included that they were within a particular age group (24 to 39 months at baseline, inclusive) and attending settings for a minimum of 15 hours per week. In some cases, settings distributed online survey links beyond the group of parents of eligible children, so that some responses had to be excluded from analysis and from the counts reported here. Inclusion criteria for practitioners were that they worked in the room(s) that would be involved in the intervention, or that included the equivalent age group in control settings. Control settings went ahead with business as usual through the 2020-21 school year and were then offered a version of the intervention in each LA starting in Autumn 2021 after the evaluation was complete.

Sample

In Nottinghamshire, after the process of recruitment, 20 settings participated in Autumn 2020 (11 in the Early Starter group, and 9 in the Late Starter group). In Hackney, similarly, there were 20 settings participating in Autumn 2020 (10 in the Early Starter group and 10 in the Late Starter group). There was one setting in Hackney for which no parents responded, and one other setting for which no practitioners responded in Autumn 2020; both were in the Late Starter group.

Of the settings from which data were collected in Autumn 2020, 3 formally withdrew from further participation in Nottinghamshire (all in the Early Starter group). A number of other settings did not respond in Summer 2021 without formally withdrawing.

By Summer 2021, the number of settings participating in the evaluation was substantially reduced for a variety of reasons, some explicitly pandemic-related (e.g. setting closures and increased workload as a result of measures responding to the pandemic) and others possibly pandemic-related but without explicit confirmation (i.e. non-response despite outreach from the research team). In total there were 8 settings in Nottinghamshire (3 Early Starter and 5 Late Starter) and 9 in Hackney (5 Early Starter and 4 Late Starter). Table 1 provides more detailed information on the number of practitioners and children (via parent report) participating in each LA before and after the intervention.

Allocation

We used a minimisation process (Altman & Bland, 2005) to allocate settings within each LA to either the Early Starter or Late Starter condition. This allowed us to draw on information obtained from the Speech and Language Therapy teams in each LA regarding the context and composition of each setting. The information used to inform the allocation process was slightly different in each LA because the contextual factors in each were different. There were some problems with practitioners' self-reported highest qualification in Hackney (e.g. "Level 3/4", making it unclear which was accurate), settings were quite homogeneous in their Income Deprivation Affecting

Children Index (IDACI) deciles[1] (all were within the three deciles indicating the highest levels of deprivation), settings were approximately evenly split between those that were Private and those that were Voluntary, and some settings had previous exposure to a similar intervention. In Nottinghamshire, only 3 settings were Voluntary, practitioners' highest qualification did not have the same problem as identified amongst the Hackney responses, and setting IDACI varied more widely (from Decile 1 to 10). Accordingly, in Nottinghamshire minimisation took into account the highest qualification of the practitioners in a setting as well as each setting's IDACI decile; Voluntary settings were allocated in a reasonably balanced way (2:1 across the Early and Late Starter groups) without including a Voluntary indicator in the minimisation process. In Hackney, minimisation was based on an indicator for Private vs. Voluntary status and previous exposure to a similar intervention.

The result of the minimisation allocation process yielded roughly comparable Early Starter and Late Starter groups in both LAs based on the information we had about settings. It is important to note that it was not possible to allocate based on child baseline data. The data collection process was prolonged as a result of the inability of the research team to visit settings due to the pandemic. Interventions in each LA needed to begin before it was possible to complete the collection of child data, which meant that allocation needed to happen based on data available on settings and practitioners to avoid delaying the start of both interventions.

Table 1 Information on participating settings in Nottinghamshire and Hackney by group (Early Starter and Late Starter)

		Nottinghamshire			**Hackney**		
		Early Start	*Late Start*	*Total*	*Early Start*	*Late Start*	*Total*
# participating at pre-test	*Settings*	11	9	20	10	10	20
	Practitioners	24	17	41	31	35	56
	Children	115	63	178	63	54	117
# participating at post-test	*Settings*	≤3	≥5	8	5	4	9
	Practitioners	7	12	19	7	13	20
	Children	20	25	45	10	12	22
Highest practitioner qualification*	*Level 3*	6	6	12	4	4	8
	Level 4 +	≥5	≤3	8	5	5	10
Setting IDACI decile	*3 or below*	≥4	≤3	7	7	7	14
	4 to 6	≤3	≤3	≤3	≤3	≤3	4
	7 or above	6	4	10	≤3	≤3	≤3
Setting type	*Private*	9	8	17	5	6	11
	Voluntary	≤3	≤3	≤3	5	4	9

Note: Child and practitioner counts do not reflect missing data on individual variables, so numbers vary between the frequencies in this table and some of the frequencies of children and practitioners included in analyses. Counts of 3 or fewer have been reported as "≤3" without specific numbers to avoid disclosure, and where necessary totals or corresponding counts have been accordingly similarly masked.
**Some practitioners did not respond to the item requesting information on their highest qualification, which in Hackney meant missing information for this variable for 2 settings.*

[1] *The measure of deprivation used is from the English Income Deprivation Affecting Children Index (IDACI) derived from the postcode of the setting. The lower the IDACI decile, the more disadvantaged the setting.*

Measures

Child language

The measures of child language used in this impact evaluation were based on the number of parent-reported words that their children used. At baseline, we used the Early Language Identification Measure (ELIM), a 50-word list developed by Law and colleagues (2020) and used previously in England (Public Health England, 2020).

At the end of the evaluation (Summer 2021), we used the Macarthur-Bates Communicative Development Inventory-III (CDI-III; Dale, 2007; Fenson et al., 2007), a similar list but with 100 words to account for children's age and development over the duration of the study. We were given authors' permission to use each instrument, and to adapt the ELIM and CDI-III to an online format using Qualtrics (https://www.qualtrics.com/uk/), by their lead authors. In the case of the CDI-III, we also purchased a set of materials in order to register our use of the instrument as agreed with the author in personal correspondence.

Previous studies have successfully used researcher-administered tests of child language (e.g. Bowyer-Crane et al., 2019). Our original evaluation plan included the use of more formal assessments of children's language (e.g. WellComm, 2010), either researcher- or practitioner-administered. However, as researcher visits to settings were not feasible, and practitioner-administered assessments would have increased the burden on staff in settings, this plan was adapted to focus exclusively on parent-report measures of child language.

Appendices 1 and 2 provide all items included in the Autumn 2020 and Summer 2021 parent questionnaires, including the word lists from the ELIM (used in Autumn 2020) but redacting the actual word list from the CDI-III used in Summer 2021 as this is proprietary and not the intellectual property of the authors of this report.

Child background information

In addition to asking parents to tick the words their children used, the parent questionnaire also asked about child gender, birthdate, birth weight, and pattern of attendance over the course of the year (including regular attendance per week in hours as well as whether this had changed over the year and any periods of time not attending e.g. due to lockdown). The original design called for investigation of the effects of different attendance patterns, but the small achieved sample made this impossible.

Practitioner confidence and skills

An instrument was developed to measure practitioners' self-reported confidence in supporting children's language and their skills in supporting children's language via interactions with children. Items were similar to those on the Hanen practitioners Teacher Interaction and Language Rating Scale (TILRS; Girolametto, Weitzman & Greenberg, 2000)) and a practitioner skill checklist from Homerton University Hospital and the Hackney Speech and Language Therapy service (Children's Integrated Speech and Language Therapy Service for Hackney and The City, n.d.). Appendix 3 provides all the items included in this instrument.

Practitioner background information

In addition to the indicators of practitioner confidence and interactions, the practitioner questionnaire included items asking about total years of experience and highest qualification.

Summer 2021 practitioner questionnaires also asked whether the practitioner had been in the setting when data were collected in Autumn 2020.

Setting context

The practitioner questionnaires included items asking for the setting name, whether the setting used particular programs to support language learning and if so, which programme(s), as well as which assessment frameworks they used, if any.

Observed practice

Video clips submitted by practitioners who agreed to participate in this aspect of the evaluation were coded using an instrument developed by the research team (see Appendix 4 for the video coding protocol). The instrument was designed to be closely aligned with the interventions and their aims, and to capture those strategies practitioners reported on in the questionnaires. The original intention was to include research visits to observe practice in settings and to assess setting quality through use of established instruments in order to measure the effects of the interventions' continuing professional development on practice (see e.g. Siraj, Kingston & Melhuish, 2015; Sylva, Siraj & Taggart, 2011; Wright et al., 2020). However, research visits were not possible in the context of the COVID-19 pandemic, which caused us to change the evaluation plan to focus on video clips of practice shared by practitioners pre- and post-intervention.

Analysis

Due to small sample sizes, throughout this report very small counts are masked in figures and tables (e.g. reported as less than or equal to 3 instead of 0, 1, 2 or 3) to avoid unintended disclosure of setting, practitioner or child identity. Where necessary, totals and percentages are similarly masked to avoid corresponding disclosure.

Approach to filtering

The number of practitioner records that could be matched/tracked from Autumn 2020 to Summer 2021 was quite small partly due to attrition and partly due to staffing changes. In total, there were 15 practitioners in Nottinghamshire (7 in Early Starter and 8 in Late Starter settings) and 14 practitioners in Hackney (only 3 in Early Starter and 10 in Late Starter settings) with records matched across pre- and post-intervention data. As a result, the practitioner analyses were conducted two ways; based on individual practitioner scores (which required individual matching and dropped any individuals without valid information at both pre- and post-test), and based on setting average scores (which allowed us to take into account, albeit imperfectly, as much valid information as possible at each time point dropping only those settings with no practitioners responding at post-test).

For child outcomes, again due to severe attrition and dropout between Autumn 2020 and Summer 2021, very low numbers of parents responded at post-test. In addition, as noted above, we filtered out any children whose parents had been invited by settings to respond but who did not fall within the age range specified for inclusion. The result was a total of 45 matched child records in Nottinghamshire at post-test, 22 in Hackney. As a consequence, results need to be interpreted with caution as this does not constitute a sufficiently-powered analytical sample to inform strong, generalisable conclusions. Instead, results provide indicative effects or lack thereof, and provide some indication of the utility of the measure used for child language from a feasibility standpoint. These numbers also preclude the use of sophisticated statistical techniques that might have been

appropriate for a larger analytical sample; we take, instead, a simple regression-based approach to assess progress in child language from pre- to post-test, bearing in mind that this approach would still ideally require a larger sample.

Validity and reliability of measures

We used correlations (Spearman's rho, given the ordinal level of measurement of the Likert scales in the practitioner questionnaire) to examine the extent to which the practitioner confidence and interaction measures related to other variables (years of experience, highest qualification) in ways that aligned with theory from the existing knowledge base. This provided us with a measure of construct validity based on the pre-intervention (baseline) data. We also computed Cronbach's Alpha coefficient as a measure of internal consistency (a traditional approach to assessing reliability).

We used a similar approach to assess validity of the parent-reported child language measure at baseline, looking for alignment with existing theory in terms of how results differed based on child background variables (gender, age in months, number of children in household, parents' education excluding responses of "I'm not sure"). Spearman's rho correlations were used where either variable did not display a normal distribution. For gender (collected as a binary variable, male/female) a t-test was used to assess whether scores differed significantly.

Effects of interventions

As the final analytic sample (including baseline and post-intervention data) was very small and not balanced across settings (i.e. there were different numbers of practitioners and children in this sample), it was not possible to treat the data using the traditional statistical approaches to account for the clustering of children and practitioners in settings. As a result, we use a combination of descriptive and basic inferential statistical tests (t-tests, regression) to examine differences between the Early Starter and Late Starter groups. We interpret the results with some caution as indicative rather than conclusive effects, acknowledging that in a larger and/or more balanced sample alternative statistical techniques would have been appropriate to account for clustering and to investigate additional details (e.g. differential effects across subgroups).

Impact of the COVID-19 pandemic on this research

Changes to measures

Children's language was originally intended to be measured using reliable and previously validated instruments (practitioner- and/or researcher-assessed). However, neither practitioner and researcher assessment was feasible due to COVID-19 upheaval in settings and restrictions on external visitors.

We also originally intended to visit settings to observe practice, but this aspect of the study also had to change due to COVID-19 restrictions on visitors to settings. As a practical alternative, we instead invited participating practitioners to submit brief videos of their practice, which could (for those in the Early Starter group in each LA) be videos already created for coaching discussions if these were part of the intervention design.

Attrition

The move to a parent-completed child language assessment and staff-completed practitioner rating scales led to low returns of data. Moreover, delays in the intervention timetable meant that post-

test data had to be collected late in the year, leading to low response rates during the summer holidays.

Limitations

Due to the limited sample size, we cannot make strong claims about the effects of the interventions (on practice or on child language) in Nottingham and Hackney based on our analyses. We can, however, discuss indicative effects where observed based on our small sample, and provide guidance on the feasibility of impact evaluation methods for future studies at a larger scale.

The inability to make allocations of settings to the Early and Late Starter groups based on child data, though arising from practical necessity as noted above, constitutes an important limitation. We control for child language at baseline in our analysis of the effects of interventions on child language to address this limitation to the greatest extent possible, but it is important to acknowledge that ideally allocations (and therefore the balance across Early and Late Starter groups) would be based on a combination of setting, practitioner and child baseline (pre-intervention) data.

Our inability to conduct in-person observations of practice was another important limitation. The use of video clips of practice in settings, though it was an attempt to approximate observations of practice without causing undue additional burden to participants, does not allow us to make claims about differences in practice over time or between Early Starter and Late Starter groups in each LA. This is because of both the lower number of video clip submissions in Summer 2021 (in some cases because of technical difficulties with OneDrive for Business), and the wide variations in the ways in which practitioners chose what activities to record and which children to include.

Descriptive information

In both Nottinghamshire and Hackney, the number of children successfully recruited (via parent consent) varied across settings. Figures 1A and 1B show this distribution, with Early Starter settings in red and Late Starter settings in blue. Of 178 children across 20 settings in Nottinghamshire, the number of children participating per setting ranged from 2 to 32, with a median of 7. Of 117 children across 19 settings in Hackney, the number of children participating per setting ranged from 1 to 17, with a median of 5. More detailed description of the sample of children, broken down by LA and by Early Starter vs. Late Starter group, is available in Appendix 5.

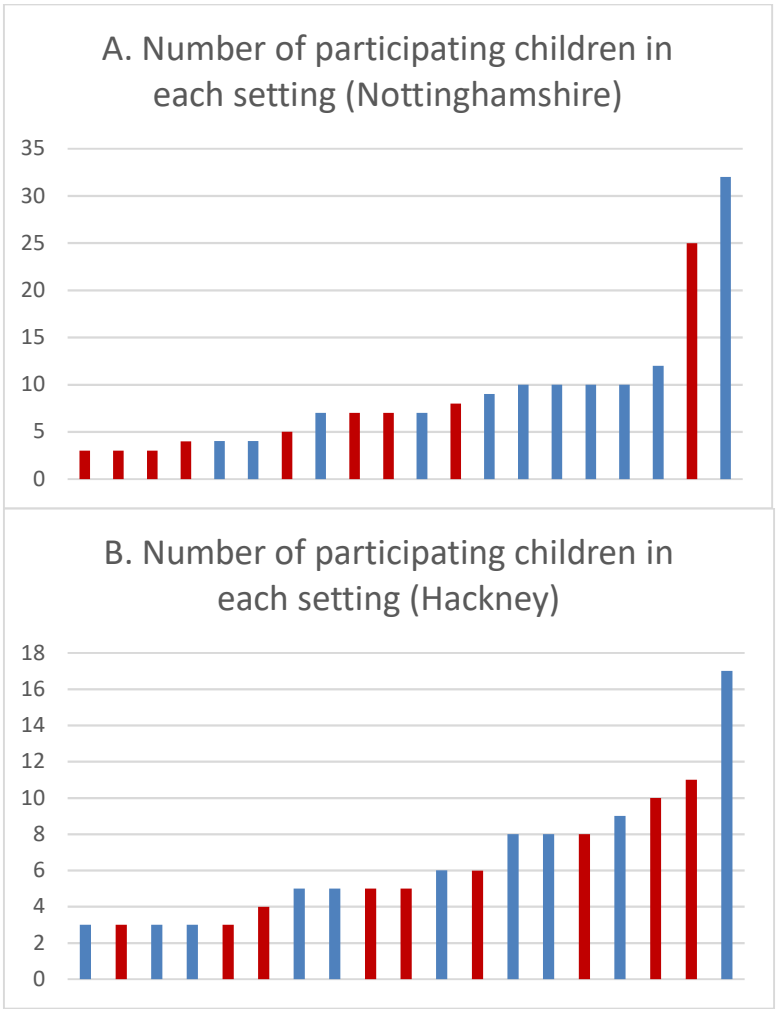

Figure 1 Distributions of participating children in each Local Authority (Nottinghamshire and Hackney)

Of 41 practitioners across 20 settings in Nottinghamshire, total years of experience ranged from 0.25 to 25 years, and the vast majority had a highest qualification of Level 3 (31 of 41). Of 56 practitioners across 20 settings in Hackney, total years of experience ranged from 0 to 39 years, and again the vast majority had Level 3 as their highest qualification (31 of 48; some practitioners did not respond to all background questions). More detailed description of the sample of participating practitioners is also available in Appendix 6.

Results

Children's language (parent-survey-based)

Validity of the instrument

At pre-test, we assessed the validity of the parent-reported children's language instrument by examining relationships with child and family characteristics to assess the extent to which any observed patterns aligned with theory and previous research (Sylva et al, 2021). It is important to note that the ELIM and CDI-III instruments have been previously validated in studies with larger samples (Law et al., 2020; Dale, Reznick & Thal, 1998; Feldman et al., 2005). There has long been evidence in the research literature to support the validity of parent-report measures of child language more broadly (Dale, 1996).

ELIM vocabulary scores at baseline were significantly and positively correlated with children's age in months (ρ=0.33, p<0.001; ρ=0.45, p<0.001 in the Nottinghamshire and Hackney samples, respectively), but not with birthweight (ρ=0.01, p=0.880 in Nottinghamshire; ρ=0.00, p=0.994 in Hackney) or number of children in the household (ρ=0.01, p=0.897 in Nottinghamshire; ρ=-0.08, p=0.373 in Hackney). Girls used significantly more words than boys in Nottinghamshire, while this difference was not significant in Hackney (t(176)=-2.33, p=0.021, d=0.36; t(115)=-0.90, p=0.369, d=0.17 in Nottinghamshire and Hackney, respectively).

Mother's education was not consistently significantly correlated with total ELIM score across LAs (ρ=0.05, p=0.550, N=163 in Nottinghamshire; p=0.227, p=0.017, N=110 in Hackney), nor was father's education (ρ=0.17, p=0.044, N=148 in Nottinghamshire; ρ=0.10, p=0.330, n=103 in Hackney); these correlations excluded the "I'm not sure" category. Children whose parents responded "I'm not sure" about the mother's highest qualification did not have significantly lower ELIM scores than those with a recorded qualification(t(176)=-1.55, p=0.122, d=-0.42 in Nottinghamshire, t(115)=-1.27, p=0.206, d=0.50 in Hackney), but children whose parents were "not sure" about the father's qualification had significantly lower ELIM scores than those with recorded father's highest qualification at least in one LA (t(176)=-1.59, p=0.115, d=0.32 in Nottinghamshire, t(113)=-2.32, p=0.022, d=0.71 in Hackney). Analysis of variance showed no significant differences in ELIM total score between children with varying use of English in the home (classified as 'English only', 'mostly English but also another language', and 'mostly a language other than English'). in Nottinghamshire (F(2, 175)=1.16, p=0.318). There was a significant difference in Hackney (F(2, 115)=14.03, p<0.01). This likely reflects differences in composition across the LAs, where Hackney had much higher numbers of children with another language spoken in the home, rather than an inconsistency in the instrument.

Where these results are somewhat inconsistent, they likely reflect compositional differences across LAs (see Appendix 5 for descriptive statistics on child background at baseline) as well as small sample size overall, but the consistent positive correlation with child age is promising in terms of the validity of parent-reported child vocabulary, as we would expect older children to use a greater number of words. Further validation of the ELIM's use in this evaluation hinges the fact that correlations between scores and both child age and birthweight were similar in our small sample to those in the larger, national sample reported by Law et al. (2020).

In addition to the above calculations, the age of the children included in this study drove the choice of instrument, given the paucity of child language measures that have been previously validated for children under 3 years old. The ELIM was appropriate for children over the age of 20 months, which made it a relevant and appropriate choice for this study. Moreover, the ELIM was especially suitable for children with English as an additional language because children were scored as using a word if they spoke the word in a home language. When it became clear that the pandemic precluded the use of a researcher-assessed measure of child language in settings, it was also important to have comparable measures of child language before and after the intervention, and the ELIM and CDI-III fit with this requirement (both involving parent report via a list of words to choose from to reflect their child's expressive vocabulary).

Early vs. Late Starter group comparison by Summer 2021

In Nottinghamshire, there was a borderline significant difference (considering a level of significance of α=0.1, given the small sample size) between the Early Starter and Late Starter groups with the

parents of children in the Early Starter group reporting the use of about 11 more words than those in the Late Starter group in Summer 2021 after controlling for children's baseline number of words as reported by parents in Autumn 2020 (Table 2; Figure 2). In other words, these results suggest that the intervention can have an effect on children's language, but that because of the small sample in this feasibility study, further research at a larger scale is needed to more securely establish a precise and reliable measure of the size of the effect of the intervention. Including other child variables (birthweight, age in months, gender) in the analysis led to negligible or no change in the intervention effect.

In Hackney there was no significant difference between Early Starter and Late Starter groups (Table 2; Figure 3).

Table 2 Child language results: Early vs. Late Starter groups, controlling for child language at baseline

Nottinghamshire

	B	*Standard Error*	*β*	*t*	*p*
Intercept	19.40	8.59	--	2.26	0.029
ELIM score	1.49	0.29	0.61	5.07	0.000
Early Starter	11.00	6.86	0.19	1.60	0.116

Hackney

	B	*Standard Error*	*β*	*t*	*p*
Intercept	12.29	15.62		0.79	0.441
ELIM score	1.46	0.39	0.66	3.70	0.002
Early Starter	3.58	9.07	0.07	0.39	0.697

Note: Results are based on small matched samples of child records with valid parent responses in Autumn 2020 and Summer 2021; N=45 in Nottinghamshire and 22 in Hackney.

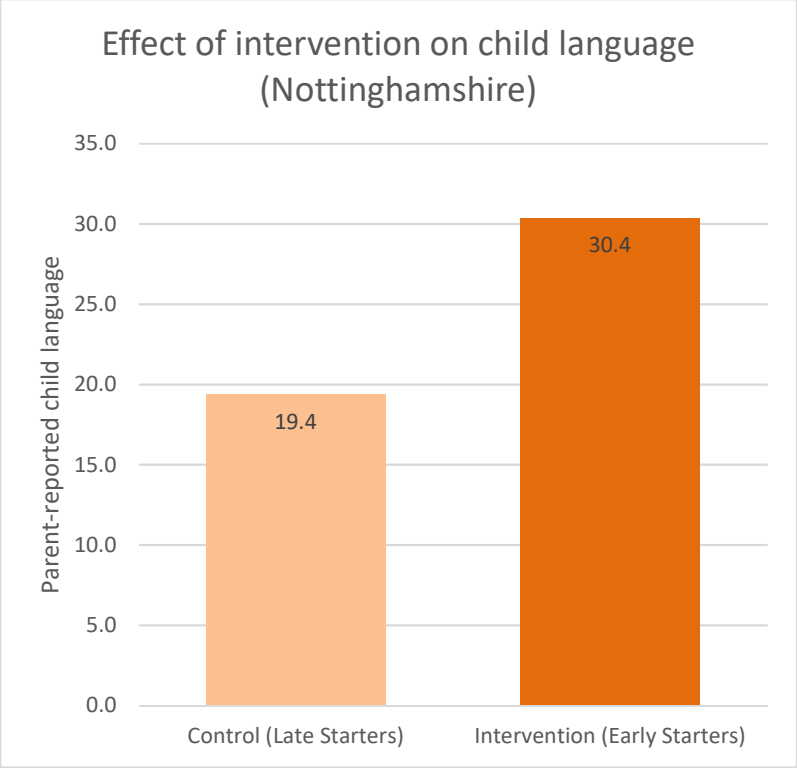

Figure 2 Effect of intervention on child language in Nottinghamshire

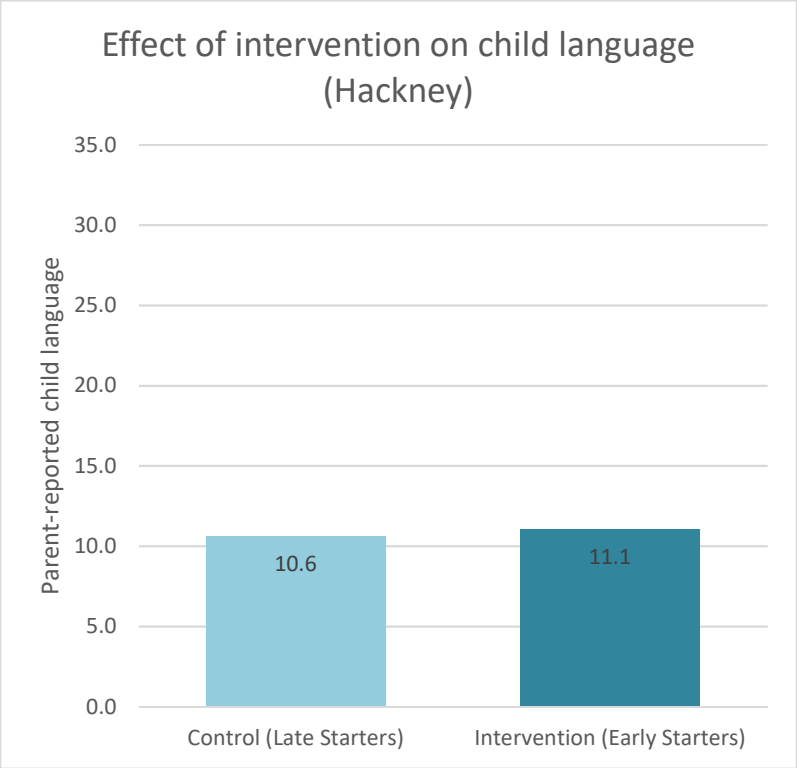

Figure 3 (Lack of) effect of child language on intervention in Hackney

Practitioners' confidence and interactions (survey-based)

Validity of the instrument

Previous research supports the validity of practitioners' own reports of their pedagogical knowledge and skills (Mathers, 2021), which underpinned our decision to use a self-report instrument to collect information about practitioners in this study. To examine the validity of our specific instrument, we investigated the extent to which practitioner-reported confidence and their skills (i.e. interactions in support of children's language development) were related to other variables including their years of experience and highest qualification.

The total score for practitioner confidence was significantly and moderately correlated with practitioner years of experience as reported in Autumn 2020 in Hackney (ρ=0.42, p=0.003). In Nottinghamshire, the same relationship was borderline significant (ρ=0.31, p=0.052). The total interaction score was not significantly correlated with years of experience in either LA (ρ=0.01, p=0.938 in Nottinghamshire, ρ=0.25, p=0.088 in Hackney). Practitioners' highest qualification did not seem to be significantly correlated with either practitioners' confidence (ρ=0.203, p=0.204; ρ=-0.02, p=0.899 in Nottinghamshire and Hackney, respectively) or interaction scores (ρ=0.116, p=0.471; ρ=0.03, p=0.853 in Nottinghamshire and Hackney, respectively), which may be due to a lack of variation in highest qualification (the vast majority of practitioners in the sample reported Level 3 qualification, see Appendix 6).

Cronbach's alpha coefficients indicated high internal consistency – traditionally used as a measure of reliability – for the confidence scale (α=0.91, α=0.95; 12 items) as well as the interaction scale (α=0.91, α=0.87; 13 items) in Nottinghamshire and Hackney, respectively. Between this high internal consistency and some indication of a roughly consistent positive relationship between the confidence scale and years of experience, we have some indicative evidence of the reliability and – to some extent – validity of the practitioner self-report questionnaire despite the small sample of practitioners in each LA.

Early vs. Late Starter group comparison

There was no significant difference in practitioners' change in total confidence and skills from Autumn 2020 to Summer 2021 between the Early Starter and Late Starter groups in either LA. This was true whether the analysis included only those practitioners who responded at both time points, or average setting scores from all practitioners responding at each time point. The lack of a significant result is hardly surprising given the very small size of the samples matched from Autumn 2020 to Summer 2021 in each LA.

Table 3 Descriptive information on practitioner outcomes in Nottinghamshire and Hackney by group (Early and Late Starters)

	Early Starter			Late Starter		
	Pre-intervention Mean (SD)	Post-intervention Mean (SD)	Difference Mean (SD)	Pre-intervention Mean (SD)	Post-intervention Mean (SD)	Difference Mean (SD)
Nottinghamshire						
Confidence	43.6 (6.1)	48.4 (7.5)	5.7 (5.1)	39.2 (8.2)	44.8 (6.1)	7.5 (7.7)
Skills/interactions	54.4 (6.5)	58.7 (6.0)	3.4 (4.0)	50.9 (5.9)	55.3 (8.2)	3.5 (9.2)
Hackney						
Confidence	45.1 (9.2)	51.1 (8.6)	11 (<0.1)	46.5 (8.9)	46.7 (7.6)	2.3 (9.3)
Skills/interactions	55.2 (5.7)	56.1 (6.0)	1.5 (7.8)	55.5 (5.9)	56.5 (5.9)	3.9 (2.6)

Note: Results are based on small matched samples of practitioner records with valid responses in Autumn 2020 and Summer 2021; N=15 in Nottinghamshire and 14 in Hackney.

Figures 4 and 5 show descriptive summaries of the practitioner responses in each LA by group via boxplots (Early Starter and Late Starter) to illustrate the lack of significant differences. These figures display the setting-level results to avoid showing data points that would be potentially identifying of individual practitioners, given the small sample size. Between the small sample overall and the lower response rate in Summer 2021, these results are inconclusive; it may simply be the case that there was not enough data to detect an effect of the interventions in either LA.

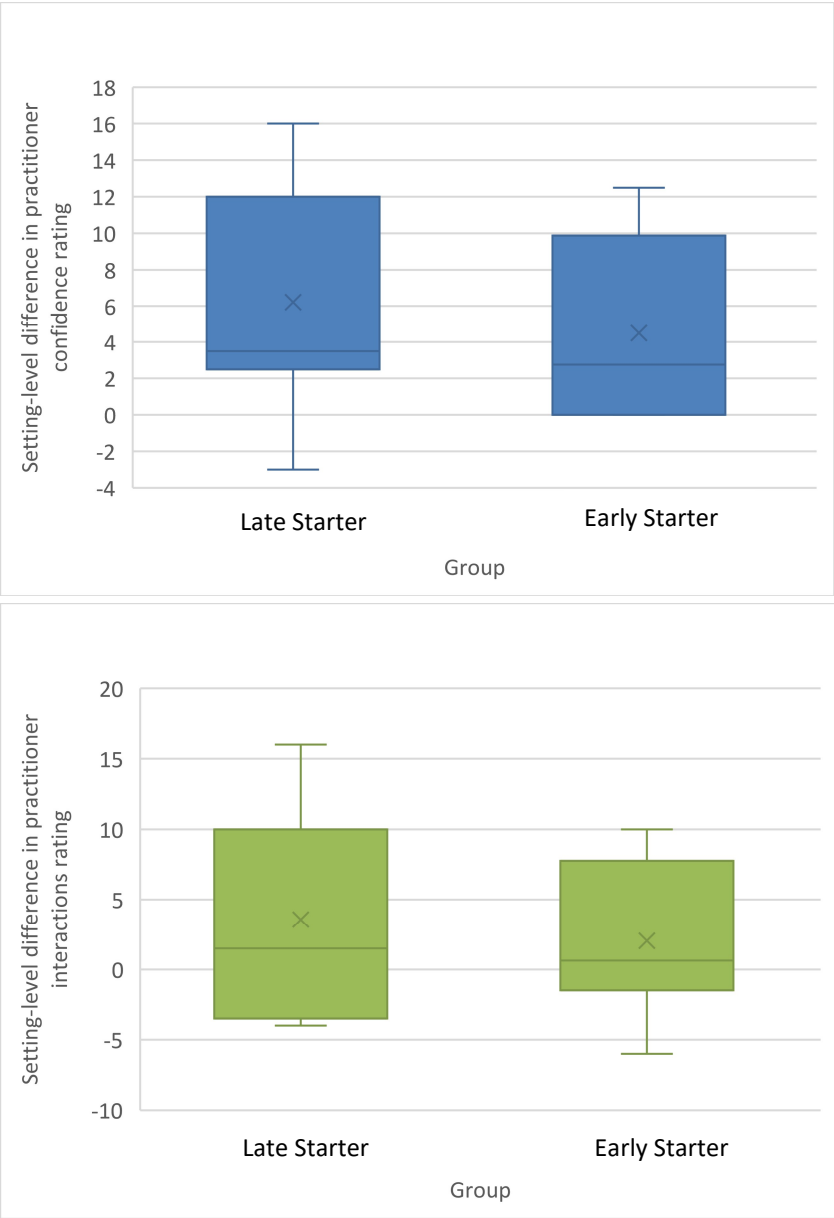

Figure 4 Practitioners' self-reported knowledge and skills -- Early Starter vs. Late Starter group comparison, Nottinghamshire

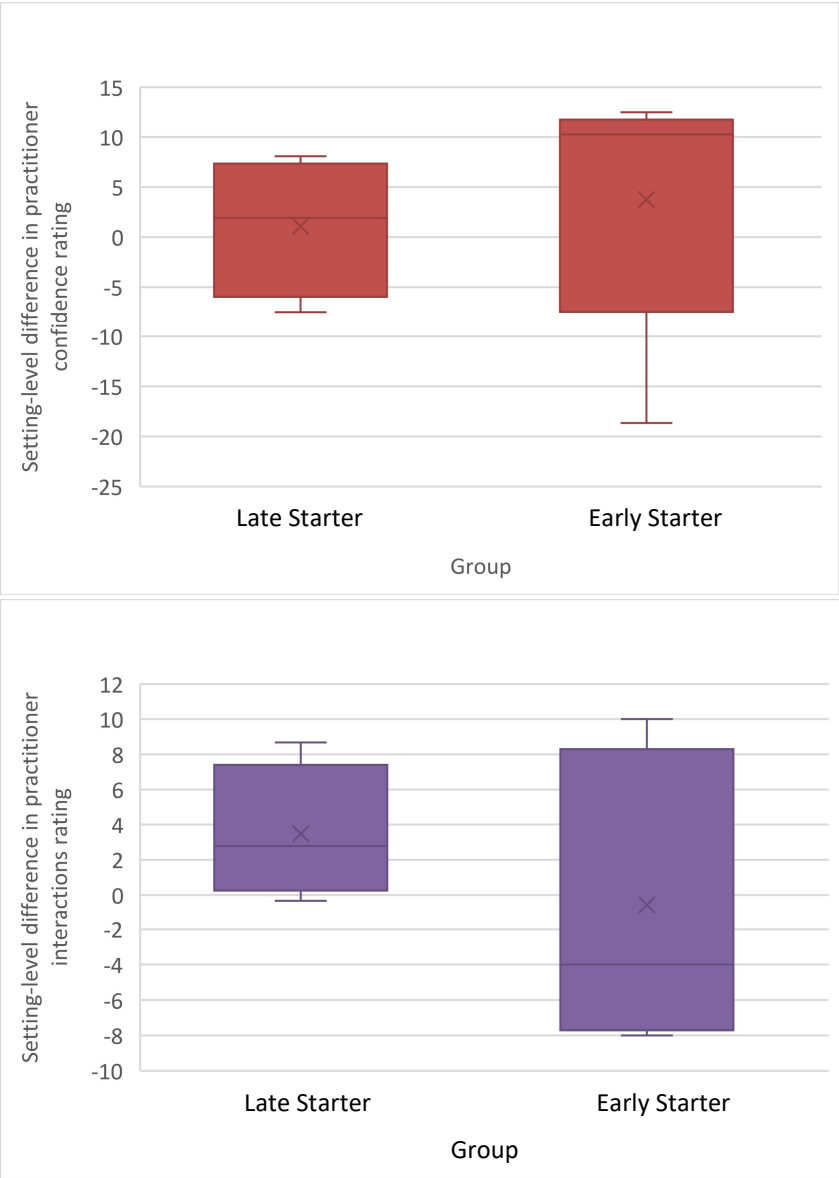

Figure 5 Practitioners' self-reported knowledge and skills -- Early Starter vs. Late Starter group comparison, Hackney

Despite the lack of significant differences, it is worth noting that there seemed to be more variation in the Early Starter group than in the Late Starter group in Hackney, but somewhat less variation in the Early Starter group than in the Late Starter group in Nottinghamshire. This might suggest some different mechanisms in the interventions in each LA that led practitioners participating in these interventions to evaluate their own practice in different ways, e.g. in Hackney with the tailored packages of support the emphasis might have been slightly different in each setting, whereas with the more standardised Nottinghamshire intervention practitioners may have become more convergent with one another in the Early Starter group in terms of their priorities, critiques and expectations of their own practice. It is also possible that this pattern reflects a difference between local authorities in terms of how many of the participating practitioners had previously engaged with similar training in the past (though this falls more within the scope of the process and implementation report by Dawson, Huxley and Garner, 2022). However, these are speculative interpretations of a result that would require further study to unpack more thoroughly. In the context of a feasibility study, a particularly salient point is that the self-report measures used for

practitioner outcomes do seem to capture variation in some meaningful way despite the lack of significant differences between Early and Late Starter groups. This suggests that the measures used may have the potential to be useful in the context of a larger-scale evaluation, though they would still benefit from further piloting if applied to evaluate a different intervention.

Observed practice (video-based)

As noted above, we originally intended to conduct research visits to observe practice in settings directly, with a researcher coding observed interactions/activities using a validated instrument. Due to the pandemic, however, it was not possible for outsiders to enter settings for health and safety reasons. As an alternative, we invited practitioners to submit short videos of their practice. These could, where applicable, be videos already created for the purpose of coaching discussions as part of the intervention in either Nottinghamshire or Hackney.

A coding instrument was developed (see Appendix 4), drawing on a combination of existing instruments, tools for observation used in the coaching process in the interventions being evaluated, and consultation with the Speech and Language Therapy teams in Nottinghamshire and Hackney to ensure relevance to the interventions. Three researchers coded a selection of three pre-intervention videos to assess the extent to which their ratings agreed, as a preliminary measure of the reliability of the instrument and of each item within it. The coding process involved attending to both the videos themselves and transcriptions of them completed by a member of the research team (e.g. to facilitate counting turns taken in conversations).

Table 3 provides calculations of the percentage of ratings on each item that agreed exactly for each video as well as the percentage of close matches (i.e. 1 point apart at most for Likert-scale items, 2 apart at most for frequency-count items). While there is no strict rule of thumb for inter-rater reliability in terms of close and exact agreement calculated in this manner, if we take 80% as an indicative threshold it is apparent that there was a high level of agreement in terms of close matches for Likert-scale items and most frequency-count items, but a few of the count items were more problematic to code, possibly related to the fact that in some instances child utterances were difficult to hear. Agreement was more difficult to achieve for coding extensions, which require the practitioner to add information or ideas (versus recasts, in which a practitioner restates or reframes what a child has said), which led to some revisions of the coding instrument.

Despite having developed a coding system that might have been sensitive enough to detect change over time and differences between the Early Starter and Late Starter groups, we focus on results with regard to feasibility. This is because the sample of videos obtained, particularly post-intervention (fewer than 10, with most concentrated in a single Late Starter setting), was not large enough to be able to make meaningful comparisons between the Early Starter and Late Starter group, nor to draw well-supported inferences about change from pre- to post-intervention. Additionally, as a result of our attempts reduce the burden on participants, we found that the videos we did receive were both so short and so varied (e.g. in terms of the numbers of children included, activities and interactions recorded) that comparisons would have been potentially problematic. Further, we found the process of obtaining video clips from practitioners proved to be a considerable burden for staff in these PVI settings. Technology was a substantial obstacle, as many PVI settings did not have access to sufficiently fast internet connections to upload videos or had tablets that allowed them to record but not to send their videos.

Table 4 Agreement on video coding items

	Item	% Exact match	% Close match
Likert-scale items	Following in with child's lead	44.4	88.9
	Joining in with child	33.3	88.9
	Encouraging child in turn-taking action	11.1	100.0
	Positioning oneself to be face to face with child	77.8	100.0
	Talking slowly enough for child to understand	55.6	100.0
	Waiting for the child to start the talking	55.6	100.0
	Asking Questions	11.1	100.0
	Providing information about the ongoing activity	33.3	100.0
	Modelling use of a variety of words through labelling	55.6	100.0
	Overall Likert-scale item agreement	42.0	97.5
Count items	Closed questions requiring yes/no answers	0.0	44.4
	Closed questions requiring short phrase answers	22.2	77.8
	Open questions	44.4	88.9
	Praise	44.4	77.8
	Repetition	11.1	33.3
	Conversational recasts	22.2	100.0
	Extension	11.1	44.4
	Expressiveness (# of child turns involving any utterance)	11.1	77.8
	Skills (# of words in the longest sentence spoken by a child)	44.4	100.0
	Overall frequency-count item agreement	25.6	74.4

Note: % match is calculated based on a selection of 3 videos each rated by 3 researchers using the coding instrument. "Likert-scale items" refer to those rated on a 1- to 5-point scale (1=(almost) not at all; 2=rarely; 3=sometimes; 4=frequently; 5=consistently). "Close match" denotes a difference of no more than 1 for Likert-scale items and no more than 2 for frequency-count items.

With feasibility in mind, then, our strong recommendations based on the experience of collecting and analysing video clips of practice, are:

- If possible, in-person research visits to observe practice should be conducted rather than analysis of video clips, particularly in the PVI sector where technology may make video-recording particularly burdensome for participants;
- Where video clips are used, clearer guidance regarding types of activities, numbers of children, etc. than we used in the present study would help to ensure a more meaningfully comparable sample of videos of practice;
- Where video clips are used, a brief questionnaire for the practitioner would be helpful to provide information about the rationales behind their selections of activities and child(ren) to record;
- Where video clips are used, the approach to transferring those videos to the research team would ideally not depend on uploading (instead using encrypted disks and posting or requiring in-person collection by a research team member);

- Researcher-recorded videos, if feasible, would facilitate consistent and comparable sampling and could be usefully combined with direct observation providing broader context regarding the interactions in settings.

Conclusions

Below, we revisit and explicitly address each of the research questions in turn, with implications for future evaluations.

What is the effect of each intervention on child speech and language?

Children in the Early Start group in Nottinghamshire used 11 more words than those in the Late Starter group on average by the end of the intervention. This difference was borderline significant.

There was no significant difference between the Early and Late Starter groups in Hackney in terms of children's language by the end of the intervention.

Given the effects of the pandemic both on the delivery of interventions (which had to rely more than originally intended on remote interaction with settings and practitioners) and on the evaluation itself (precluding research visits to observe settings and researcher assessments of child language, as well as severe attrition between baseline and post-intervention data collection), these findings do not demonstrate that the interventions are ineffective. It is possible that positive effects would have been found under more favourable conditions to allow for the implementation of the interventions *as originally planned* and impact evaluation *as originally intended*. Moreover, the small sample size and the sole reliance on parent-reported child outcome mean that definitive answers to the main research questions are not possible.

What is the effect of each intervention on practitioner outcomes (observed practice, confidence and professional knowledge)?

No significant effects on practitioners' confidence and skills were found. This result, however, relies on an extremely small sample of practitioner responses that could be matched from baseline data collection in Autumn 2020 to post-intervention data collection in Summer 2021. It is possible that the delivery of the interventions was adversely affected by the rapid shift to online activities due to the pandemic. However, it is also possible that interventions had subtle effects on practitioners' self-report at the end of the year, making those participating in interventions in both LAs more aware of, or more critical of, certain aspects of their practice. If the interventions led to deeper knowledge of effective practice, then self-critical practitioners in the intervention group might give themselves lower scores at post-test. In future large-scale evaluations, we suggest combining practitioner self-report with more objective measures of practice based on researcher observation.

How do practitioner outcomes relate to child outcomes?

Given the lack of significant effects on practitioner outcomes in this study, we stopped short of investigating relationships between practitioner-level change and child-level change. However, in a larger-scale study with more power to detect effects, it may be possible to show which practitioner thoughts and behaviours are associated with gains in children's language.

What early language child assessments are appropriate as pre- and post-test measures for a future larger-scale study?

Due to the effects of the pandemic on settings, interventions and the evaluation of them, we were limited in the extent to which we could explore this, as the assessments of child language that we hoped to use were not feasible while research visits were impossible to conduct.

What we did find was some evidence that parent assessments of child language were reasonably valid as a measure of vocabulary, in the absence of more complete (i.e. covering a broader range of language skills) and more objective measures. We recommend the use of more formal and objective assessments in a larger-scale study if possible, but we also suggest using parent-reported child vocabulary alongside those assessments as a useful tool to understand language use at home as well as in settings. Parent report on the assessments used here are particularly appropriate for children whose first language is not English because they get "credit" for words spoken in the home language.

What measure(s) of observed practice, practitioner confidence and professional knowledge are appropriate as pre- and post-test measures for a future larger-scale study?

Again, we were limited in the extent to which we could assess measures of observed practice given the limitations on our analysis imposed by the low response rate for video clips of practice at post-test. Given the technological difficulties experienced by settings, not only in making video-recordings of practice but also – and perhaps more markedly – in sharing them beyond their settings, we would recommend research visits to settings to observe practice in person over the use of video recordings where feasible in a larger-scale study. If resources for future evaluation allow, pre- and post-videos could be more structured, with practitioners asked to film specific activities with specific types of children.

We conclude by noting how much the Oxford team came to admire the competence, commitment and creative problem-solving shown by the Speech and Language Therapists and practitioners participating in the CECIL project. Obstacles were encountered at every step of intervention delivery and evaluation activity. The Speech and Language Therapists who led the interventions were not daunted by the sudden shift to online delivery nor last-minute changes in timetables brought on by staff absence or equipment failure. Practitioners were equally committed, staying late to finish a video or changing family plans to attend an online workshop. The primary goal of the Oxford impact study was to test the feasibility of a large scale RCT evaluation. To a large extent, this was accomplished. Much was learned, especially the utility of parent-reported measures and the unique needs of the PVI sector. Perhaps the most important learning of all, however, was how committed Speech and Language Therapists are to supporting settings, and how enthusiastic and competent are the practitioners who take part in intervention efforts. All of this bodes well for future intervention and evaluation.

References

Altman, D. G. & Bland, J. M. (2005). Treatment allocation by minimisation. *BMJ*, *330,* 843. doi:10.1136/bmj.330.7495.843

Bowyer-Crane, C., Nielsen, D., Bryant, M., Dharni, N., Heald, R., Storr, C., & Dickerson, J. (2019). A randomised controlled feasibility trial and qualitative evaluation of an early years language development intervention: study protocol of the "outcomes of Talking Together evaluation and results" (oTTer) project. *Pilot and Feasibility Studies*, *5*(1). https://doi.org/10.1186/s40814-019-0498-2

Dale, P.S. (2007). *The MacArthur-Bates Communicative Development Inventory – III.* Baltimore, MD: Brookes Publishing.

Dale, P. S. (1996). Parent report assessment of language and communication. In K. N. Cole, P. S. Dale, & D. J. Thal (Eds.), *Assessment of communication and language,* Vol. 6, pp. 161–182). Baltimore, MD: Brookes Publishing.

Dale, P. S., Reznick, J. S. and Thal, D. J. (1998) A parent report measure of language development for three-year-olds. International Conference on Infant Studies, Atlanta, Georgia.

Dawson, A., Huxley, C. & Garner, O. (2022). *Coaching Early Conversation Interaction and Language (CECIL) Evaluation: Implementation and Process Evaluation.* Brighton, UK: Institute for Employment Studies (IES).

Dearden, L., Sibieta, L., & Sylva, K. (2011). The socio-economic gradient in early child outcomes: evidence from the Millennium Cohort Study. *Longitudinal and life course studies, 2*(1), 19-40.

Dockrell, J. E., Stuart, M., & King, D. (2010). Supporting early oral language skills for English language learners in inner city preschool provision. *British Journal of Educational Psychology*, 80(4), 497-515.

Duncan, G. J., Magnuson, K. & Votruba-Drzal, E. (2017). Moving beyond correlations in assessing the consequences of poverty. *Annual review of psychology, 68*(1), 413-434.Feldman, H. M., Dale, P. S., Campbell, T. F., Colborn, D. K., Kurs-Lasky, M., Rockette, H. E., & Paradise, J. L. (2005). Concurrent and Predictive Validity of Parent Reports of Child Language at Ages 2 and 3 Years. *Child Development*, *76*(4), 856–868. https://doi.org/10.1111/j.1467-8624.2005.00882.x

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., & Reznick, J. S. (2007). *MacArthur-bates communicative development inventories*. Baltimore, MD: Brookes Publishing.

Fricke, S., Bowyer-Crane, C., Haley, A. J., Hulme, C., & Snowling, M. J. (2013). Efficacy of language intervention in the early years. *Journal of Child Psychology and Psychiatry, 54*(3), 280–290. https://doi.org/10.1111/jcpp.12010

Children's Integrated Speech and Language Therapy Service for Hackney and The City (n.d.). *Launchpad for Language*. Retrieved from https://gethackneytalking.co.uk/educational-professionals/launchpad-for-language/

Girolametto, L., Weitzman, E., & Greenberg, J. (2000). *Teacher interaction and language rating scale*. Toronto, Ontario: Hanen Early Language Program.

Law, J., Charlton, J., McKean, C., Watson, R., Roulstone, S., Holme, C., Gilroy, V., Wilson, P., Rush, R. (2020). *Identifying and Supporting Children's Early Language Needs.* Newcastle, UK: Newcastle University.

McDonald, D., Proctor, P., Gill, W., Heaven, S., Marr, J., & Young, J. (2015). Increasing early childhood educators' use of communication-facilitating and language-modelling strategies: Brief speech and language therapy training. *Child Language Teaching and Therapy*, 31(3), 305-322.

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. https://doi.org/10.1016/j.ecresq.2020.10.010

Melhuish, E. C., & Gardiner, J. (2018). *Study of Early Education and Development (SEED): Impact study on early education use and child outcomes up to age four years [Research Brief]*. London: Department for Education.

Public Health England (2020). *Early language identification measure and intervention: Guidance handbook.* London: Public Health England, Department of Health & Social Care, Department for Education.

Sibieta, L., Kotecha, M., & Skipp, A. (2016). *Nuffield Early Language Intervention: Evaluation Report and Executive Summary*. London: Education Endowment Foundation.

Siraj, I., Kingston, D., & Melhuish, E. (2015). *Assessing quality in early childhood education and care. Sustained shared thinking and emotional wellbeing (SSTEW) Scale for 2–5 year olds provision*. London: UCL and IOE Press.

Sylva, K., Joseph, A., Ereky-Stevens, K & Lindorff, A (2021) Responding to covid requirements in an early childhood language intervention RCT. European Early Childhood Education Research Association online conference. Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2004). The final report: Effective pre-school education. London: Institute of Education.

Sylva, K., Siraj, I. & Taggart, B. (2011). *ECERS-E : the four curricular subscales extension to the early childhood environment rating scale (ECERS-R)*. New York, NY: Teachers College Press.

Weitzman., E & Greenberg, J. (2002) *Learning Language and Loving It: A guide to promoting children's social, language, and literacy development in early childhood settings*. Toronto: The Hanen Centre.

WellComm. (2010). *A Speech and Language Toolkit for Screening and Intervention in the Early Years*. GL Assessment. Retrieved from: https://www.gl-assessment.co.uk/assessments/products/wellcomm/

Wright, H., Carr, D., Wiese, J., Stokes, L., Runge, J., Dorsett, R., Heal, J., & Anders, J. (2020) *URLEY Evaluation Report*. London: Education Endowment Foundation. Accessible from: https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/URLEY_Report.pdf

Appendix 1: Parent questionnaire (Autumn 2020)

This parent questionnaire was adapted from the ELIM instrument by Law and colleagues (2020).

PARENT QUESTIONNAIRE

This short questionnaire is designed to find out about your child and their language development. It should not take more than 10-20 minutes to complete. We will ask the same questions again at the end of the evaluation to see whether and how your child's word use has changed.

Your answers will be kept confidential and used only for research purposes by the Oxford research team. Your child's name will never be used in any report of results. The information sheet and consent form from the University of Oxford are attached to this document.

In this first section, we'll ask a bit about your child and family. There are no "right" or "wrong", "good" or "bad" answers.

1. What is **your child's name**? (First Last, e.g. Helen Smith)	
2. What is **your child's gender**? (Circle your answer)	Male Female
3. What is **your child's date of birth**? (MM/DD/YYYY)	
4. What was **your child's birthweight**, if known? (e.g. 3.54kg)	
5. **How many children are there in your household**, including the child participating in this evaluation? (Circle your answer)	1 2 3 4+
6. What is the child's **mother's highest educational qualification**? (Tick one)	☐ GCSE or below ☐ A-level ☐ Undergraduate ☐ Postgraduate or higher ☐ I'm not sure
7. What is the child's **father's highest educational qualification**? (Tick one)	☐ GCSE or below ☐ A-level ☐ Undergraduate ☐ Postgraduate or higher ☐ I'm not sure
8. What **languages are spoken at home**? (Tick one)	☐ English ONLY ☐ Mostly English, but also other language(s) ☐ Mostly a language other than English

Please turn over!

In this section, we'll ask about what words your child says, in English <u>OR</u> in any other home language. Please choose all of the words that your child says in each list.

There are no "right" or "wrong", "good" or "bad" answers. Please answer as honestly as possible, as we are keen to learn about your child's use of language.

Which words does your child say? Please take a look at the words below and tick which words you have heard your child say.										
1. Mummy/mum		11. Aeroplane		21. Towel		31. Fit		41. Wet		
2. Bye/bye bye		12. Car		22. Bed		32. Like		42. After		
3. No		13. Book		23. Settee/sofa		33. Rip/tear		43. Day		
4. Ball		14. Milk		24. School		34. Shake		44. This		
5. Juice		15. Hat		25. Friend		35. Think		45. Our		
6. Owch/ow		16. Shoe		26. Person		36. Gentle		46. Where		
7. Cat		17. Leg		27. Hello/hi		37. Fast		47. All		
8. Thank you		18. Pillow		28. Shopping		38. Happy		48. Much		
9. Cold		19. Rubbish		29. Carry		39. Last		49. Need to		
10. Hug/cuddle		20. Plate		30. Finish		40. Tiny		50. If		

When finished, please return the consent form and questionnaire to your child's playgroup/nursery. If you prefer, you may instead scan or take a photo of each page & return to the research team at early.language@education.ox.ac.uk**.**

Appendix 2: Parent questionnaire (Summer 2021)

PARENT QUESTIONNAIRE

This short questionnaire is designed to find out about your child and their language development. It should not take more than 10-20 minutes to complete.

Your answers will be kept confidential and used only for research purposes by the Oxford research team. Your child's name will never be used in any report of results. The information sheet and consent form from the University of Oxford are attached to this document.

In this first section, we'll ask a bit about your child and family. There are no "right" or "wrong", "good" or "bad" answers.

9. What is **your child's name**? (First Last, e.g. Helen Smith)	
10. What is the name of **your child's preschool**?	
11. What is **your child's gender**? (Circle your answer)	Male Female
12. What is **your child's date of birth**? (DD/MM/YYYY)	
13. What was **your child's birthweight**, if known? (e.g. 3.54kg)	
14. **How many children are there in your household**, including the child participating in this evaluation? (Circle your answer)	1 2 3 4+
15. What is the child's **mother's highest educational qualification**? (Tick one)	☐ GCSE or below ☐ A-level ☐ Undergraduate ☐ Postgraduate or higher ☐ I'm not sure
16. What is the child's **father's highest educational qualification**? (Tick one)	☐ GCSE or below ☐ A-level ☐ Undergraduate ☐ Postgraduate or higher ☐ I'm not sure
17. What **languages are spoken at home**? (Tick one)	☐ English ONLY ☐ Mostly English, but also other language(s) ☐ Mostly a language other than English

Please turn over!

In this section, we'll ask about what words your child says, in English OR in any other home language. Please choose all of the words that your child says in each list.
There are no "right" or "wrong", "good" or "bad" answers. Please answer as honestly as possible, as we are keen to learn about your child's use of language.
Please note: if your child is not talking yet, or if s/he is talking, but you can not understand, please tick this box ☐

[Word list redacted as the CDI-III is a proprietary instrument and not the intellectual property of the authors of this report.
Words were listed in four columns of 25 each, with a box to tick accompanying each word.]

When finished, please return the consent form and questionnaire to your child's preschool. If you like, you may instead scan or take a photo of each page & return to the research team at early.language@education.ox.ac.uk.

Appendix 3: Practitioner questionnaire

PRACTITIONER CONFIDENCE AND SKILLS QUESTIONNAIRE

This short survey is designed to find out how confident you feel now as an early years practitioner. It should only take around 10 minutes to complete.

Your answers will be kept confidential and used only for research purposes by the Oxford and IES research teams. Your name and your setting's name will never be used in any report of results. The information sheet and consent form from the University of Oxford are attached to this document, and the IES privacy policy can be found at: https://www.employment-studies.co.uk/cecil-privacy-policy

The first set of questions are designed to find out **how confident you feel as an early years practitioner aiming to support children's language development**. Most of the statements use wording like this: **"How confident are you that …?'** We want to find out the extent to which you feel you have the knowledge/skills needed to produce a specific outcome (e.g. increased children's vocabulary).

There are no right or wrong answers and practitioners vary a great deal in their confidence. Please be as honest as you can, because we are keen to learn about staff confidence across a broad range of skills.

How confident are you in your knowledge and skill at each of the following:	Not at all	Slightly	Somewhat	Fairly	Very much
18. Helping **typically developing** children make good progress in their language skills	1	2	3	4	5
19. Helping **children with language delay** make good progress in their language skills	1	2	3	4	5
20. Helping **children with EAL** make good progress in their language skills	1	2	3	4	5
21. Crafting good **questions** for your children	1	2	3	4	5
22. Enabling children to **ask their own questions**	1	2	3	4	5
23. Supporting children to be **good listeners**	1	2	3	4	5
24. **Suggesting activities that families can do** to support children's language development	1	2	3	4	5
25. Supporting children to **be confident** in communicating their wishes and ideas	1	2	3	4	5
9. **Motivating children** to want to communicate more with peers and adults	1	2	3	4	5
10. **Assessing children's language** to identify their need for support	1	2	3	4	5
11. **Engaging other early years staff in changes** to language practice	1	2	3	4	5
12. **Making referrals** for extra support for a child with language difficulties	1	2	3	4	5

Please turn over!

This next section asks about **how you interact with individual children**. Please rate the extent to which you do the following things, answering as honestly as you can.

What you normally do when interacting with a child or group?	Never	Rarely	Sometimes	Often	Very often
1. Wait for child to start the talking – with words, sounds, gestures or looks	1	2	3	4	5
2. Follow child's lead in play	1	2	3	4	5
3. Join in with child play	1	2	3	4	5
4. Position yourself to be face to face with child	1	2	3	4	5
5. Use a wide range of questions	1	2	3	4	5
6. Encourage child in turn-taking	1	2	3	4	5
7. Imitate what child has said or done	1	2	3	4	5
8. Comment on what child is doing	1	2	3	4	5
9. Repeat what child has said, using the correct form of speech	1	2	3	4	5
10. Extend what child has said, e.g. by linking to a previous event or providing another example	1	2	3	4	5
11. Model a wide variety of words, e.g. adjectives and connectives such as 'because'	1	2	3	4	5
12. Praise child	1	2	3	4	5
13. Talk slowly enough for child to understand	1	2	3	4	5

This final section asks **about you and your setting**. Your name and setting's name are needed to help us match your Autumn and Summer answers. It will not be used in any other way, and once matched, your name will be deleted from all records.

1. Your name	
2. Highest professional qualification (circle one)	Level 2 childcare Level 3 or 4 EYT QTS Other (please specify):
3. Does your setting assess language skills? If so, what measure do you use? (e.g., EYFS Development Matters, WellComm Screen)	
4. Since September 2020, what training have you had to support children's language?	Let's Interact Other (please specify):
5. For **how many years** have you been working (not including training/apprenticeship) as an Early Years professional?	
6. Your setting's name	
7. Date of completion (DD/MM/YYYY)	

When finished, please put the consent form and questionnaire in the response envelope given to your setting. If you prefer, you may instead scan or take a photo of each page & return to the research team at early.language@education.ox.ac.uk Appendix 4: Child descriptive information

Appendix 4: Video coding protocol

VIDEO-CODING PRACTITIONER JOINT ENGAGEMENT AND LANGUAGE INTERACTION WITH TWO- AND THREE-YEAR-OLD CHILDREN IN EARLY YEARS PROVISION

Description of the video-recordings:

Practitioners were asked to make, and share, with the researchers a 2–3-minute video (at least 2, but no more than 5 minutes) of them playing with a child (or a group of children) in an informal setting. The instructions asked practitioners to choose a quiet location with not too much background noise, to arrange for someone else to make the recording, if possible, and to aim for an interaction that is not too structured, or too adult led. They were reminded that the aim for researchers was not to see practitioners to 'make the children talk', but to reflect on the strategies practitioners are using during the interaction to help children join in and interact and communicate in the widest sense (not just talking). Video-clips were transcribed.

Most video-taped situations involve children and practitioners interacting with objects – building something, creative activities (drawing) or pretend play (e.g. with dolls), but also exploring objects and describing features. Most videos we received show one practitioner interacting wither with one or two children, and sometimes with a small group of children. Group activities are sometimes more structured, with children sitting in a half-circle and the practitioner leading an activity that asks children to join in, one after the other. The age of the children and their developmental (language) stage varies.

Instructions for use:
Procedure
The coding system is designed to evaluate practitioners'' interaction with two- and three-year old children in early years provisions, with a focus on the strategies practitioners use to help children join in and interact and communicate. While there is a clear focus on verbal exchanges, the coding system intends to also capture interactional exchanges in a wider sense – considering for example also joint engagement in play, with turn-taking actions between practitioner and child. The coding system consists mainly of rating scales, but also includes some event codes.

Please read over the instructions carefully and familiarise yourself with the scale. Watch the video-clip and read the transcript at least once all the way through without stopping. Take some notes (short comments) related to the items to code and briefly reflect on your notes. Watch the video and read the script a second time (and if needed more often, with pauses for reflection) and complete the ratings and event coding.

Rating scales
Items are rated on a 5-point scale. A rating of 1 indicates that the practitioner almost never uses a strategy, and a rating of 5 indicates that the strategy is used consistently. Ratings of 2 and 3 indicate that practitioners could clearly improve on the strategies they are using, and a rating of 4 and 5 indicate that strategies are applied appropriately and frequently (or consistently) throughout the interaction. Give a rating of 1 if you cannot observe a strategy at all or when there is only some very weak indication that a strategy is applied during some very rare instances. Give a rating of 2 if there are clear signs that the strategy is applied appropriately at least during some (infrequent) instances. Give a rating of 3 if a strategy is applied appropriately at least sometimes during the interaction, but there is clearly room for improvement. Give a rating of 4 if a strategy is applied competently and frequently, and a rating of 5 if the strategy is applied consistently and competently consistently throughout the interaction.

Frequency (event) recording
For some of the strategies, practitioner verbalisations will be highlighted, labelled and counted. For all event codes, please highlight and label the particular verbalisation in the transcript, and label it. We allow for double coding.

Collecting information on children's verbal engagement and skills
We take a note of the number of words in the longest word sentence produced by the child (one of the children) during the recorded interaction.

We take a note of the number of 'conversational turns' children take during the recorded interaction. Any initiation of a child verbalisation counts, including vocal sounds, and utterances labelled as inaudible in the transcript).

CODING SYSTEM

Facilitating joint engagement in play

Following in with child's lead: (almost) not at all – rarely – sometimes – frequently – consistently
Following the child's lead involves practitioners observing and noticing what children are interested in, waiting to give them a chance to initiate activities or get involved, and expressing the expectation for the child to act, listening to what the child is expressing, and responding with enthusiasm and interest to the child's initiations of activities (Girolametto, Weitzman, & Greenberg, 2000). Practitioners go along with children's activities, they stay focussed, and avoid distracting the child. They express enjoyment and interest in the activity (positive emotional tone, comments on what they are both doing). If the child changes an activity, the practitioner moves along with the child and shifts focus. When the child initiates (verbally or non-verbally), practitioner follows the lead, by responding verbally to initiations and by using animations, and by avoiding commands and vague acknowledgements (uh huh; yes, that's right) (see Girolametto et al., 2000).
Joining in with child: (almost) not at all – rarely – sometimes – frequently – consistently
If a child is playing, the practitioner is actively joining in as a play partner, and engages with the child following their directions, and tuning into their ideas and expressions. The practitioner is allowing the child to steer the play, and offers ideas that are related to the play, without taking over/dominating (see Girolametto et al., 2000).

Encouraging child in turn-taking action: (almost) not at all – rarely – sometimes – frequently – consistently
Sometimes observing, waiting and listening to a child isn't enough to encourage a child to initiate. If needed, and to enable joint engagement in play, practitioners provide some inputs to encourage the child to get involved. They set up a situation, and wait to see if the child responds, and repeat this to see if they can encourage involvement and child initiation (Hanen). For example, the practitioner pauses after initiating an action, and creates a space that suggests there is an expectation for the child to respond in action (taking turns rolling a ball; taking turns to build a tower). Or the practitioner makes an utterance that encourage turn-taking actions ('Would you like to put on the next block?' 'Yummy tea! Would you like to have a try?') and responds with animation If the child gets involved (see Girolametto et al., 2000). *Note: The focus here is on turn-taking actions, rather than conversations.*

Making sure the child understands verbal inputs

Positioning oneself to be face to face with child: (almost) not at all – rarely – sometimes – frequently – consistently
Practitioner positions himself/herself to be at the child's level whenever possible. This involves bending down to the child be close, adjusting her/his physical level by sitting on the floor or a child-sized chair, leaning forward to facilitate face-to-face interaction (see Girolametto et al., 2000)

Talking slowly enough for child to understand: (almost) not at all – rarely – sometimes – frequently – consistently
Practitioner speaks at a speed that is in tune with the child's ability and engagement and allows for the child to understand verbal input.

Encouraging turn-taking conversation

Waiting for the child to start the talking: (almost) not at all – rarely – sometimes – frequently – consistently
Practitioner pace of verbal initiations is slow enough (in tune with the child's ability and engagement) to allow for the child to join in, and to respond verbally. During the interaction, the practitioner expresses the expectation for the child to join in the conversation – signalling this expectation through with words, sounds, gestures, or looks. Practitioner gives children a chance to talk, by pausing and leaving sufficient time for a child to respond verbally. When the child is talking, the practitioner allows children to complete their messages.

Asking questions
Using a variety of questions to elicit a verbal response from the child: (almost) not at all – rarely – sometimes – frequently – consistently
Practitioners invite the child to talk through asking questions and waiting expectantly for children to respond to questions. However, importantly, here practitioners do not only use yes/no questions but only use those to obtain information and to clarify, and they avoid test- or rhetorical questions. WH questions are used to facilitate and extend back and forth conversations (see Girolametto et al., 2000).

Asking child a question: frequency (event) recording
Practitioner asks child a question. Total number of questions will be counted (and we will make no distinction between rhetorical questions, or questions that are meant as suggestions, and 'real questions that intend to facilitate conversational exchanges). Each question will then be assigned to one of the following categories: a) closed questions that require only a yes/no answer; b) closed question requiring single-word answers (that are not yes/no answers) or short phrase answer (e.g. What have you got there? – A ball!; Where is the dog? – In the house!), c) open questions that initiate the child to describe or explain something, to think and reflect, to imagine. This can be a why or what question, or a how (e.g. How does it work? Why is the baby hungry? Where is the train going today?).

Modelling use of rich language

Providing information about the ongoing activity: (almost) not at all – rarely – sometimes – frequently – consistently
Practitioner comments on what they or the children are currently actively engaged with. For example, they describe what they or the children are doing, or refer to objects they are engaged with. This includes statements of questions like: 'That's the spoon you are holding', 'How many blocks have you got there?'. The information the practitioner provides related to the ongoing activity can also add additional information, and thus serve to extend beyond statements and questions that

focus on observable actions and objects. Practitioners could for example comment and use questions to inform, project, pretend/imagine, explain, and talk about feelings or the future (see Girolametto et al., 2000). E.g. I have got some big blocks here, let's see if I can stack them!'; 'Look at the dolly – she is saying, please can you feed me? I think she is hungry, and she wants a cuddle!' Importantly, this includes questions and statements that focus on the internal states, e.g. 'You think that is funny', 'You like rolling the ball!'

Modelling use of a variety of words through labelling: (almost) not at all – rarely – sometimes – frequently – consistently
Practitioner uses a variety of vocabulary (nouns, verbs, adjectives, adverbs) during the interaction. They emphasise key words, repeat words, label objects/attributes/events, and avoiding non-specific words (e.g. it, this, that, there, thank you), and adjust complexity of vocabulary for different children in the group (see Girolametto et al., 2000).

Responding to the child's vocalisations and verbal expression

Praise: frequency (event) recording
Praising child's verbal expressions.

Repetition/Imitation: frequency (event) recording
Imitating child vocalisation/verbal expression/repeating what the child has said.
Conversational recasts: frequency (event) recording
A conversational recast is an (immediate) response to a child's utterance in which the adult repeats some or all of the child's words and adds new information while maintaining the meaning expressed by the child. The additional information added by the adult response can be syntactic (adding new or different grammar), semantic (adding word meaning information) or phonological (changing a sound) (see Cleave et al., 2015). In a recast, a child's verbalisation gets expanded into a different type or more elaborated sentence.

A recast can occur when the caregiver changes the voice or modality of the child's utterance. For example, the practitioner repeats what the child has said but changes perspective in doing so (e.g. child: I'm gonna build my house! Practitioner: You're gonna build your house!) we count this as a recast. A statement (child: Him need juice) can also be recast as a question (practitioner: Does he need some juice?) – other examples include: 'Doggy house!' – 'Is the doggy in the house?', Isn't the doggy in the house?' or 'The doggy is in the house, isn't it?'.

Expansions are a type of recast, where they adult utterance maintains the child words and basic meaning but modifies the child's sentence by changing structural or semantic details without changing the sentence modality. E.g. 'Him need juice' – 'He needs juice' (see Claeve et al., 2015). Recasts can be corrective, fixing an error in the child utterance, but does not have to be corrective. An expansion can add new words to what the child is saying. E.g. 'Car!' – 'It's a car' or 'Doggy!' – 'It's a big, black dog!' or 'Want car!' – 'You want another car?'

Practitioners often facilitate conversations and turn-taking interactions in asking follow-up questions: 'Doggy!' – 'Shall we put the doggy in the house?'. We do not count these questions as an expansion, because it is not clear if the child's basic meaning is kept or if the practitioner is making a play suggestion.

4 Extension: frequency (event) recording
Semantic extensions are distinguished from a recast/an expansion. In an extension, the caregiver response to the child's verbalisation continues the child's topic and adds new information. In

contrast to an expansion, an extension does not necessarily contain any of the child's words and can change the sentence modality. Child: Dog running fast! Practitioner: He's in a hurry! Or 'Baby cry' – 'The baby is sad'. Extensions are often used in conjunctions with expansions (which are a form of recast). E.g. 'Doggy house!' – 'It's the dog's house!' [expansion] 'He is a large dog!' [extension] (see Claeve et al., 2015).

Extending questions
Extensions can also be questions, that follow up on the child's utterance and serve to extend the information. E.g.
'They got squished!' – 'Oh no, who squished them?!'
'The sheep are running away!' – 'Who is chasing them?!'
'The baby is crying!' – 'Why is the baby crying?!'
'The train is leaving the station!' – 'Where is it going?!'

'The baby is crying!' – 'Is the baby hungry?'
'The train is leaving the station!' – 'Is it making a lot of noise?'
'The sheep are running away!' – 'Is the wolf going to get them?'
'I found a car!' – 'What colour is the car?'

Practitioners often facilitate conversations and turn-taking interactions in asking follow-up questions:
'Doggy!' – 'Shall we put the doggy in the house?' or' What do you want to do with the doggy?'
We do not count these questions as extensions, because they are not linguistic extensions per se, but serve to clarify, or make suggestions.

Child verbal expressiveness and skills

Expressiveness: Count the number of turns during the recorded interaction where a child produces one or more utterances (including any type of verbal expression, e.g. inaudible words or statements, or vocal signals, but not just the child laughing).

Skills: Longest word sentence produced – counted in number of words.

References:

Cleave, P. L., Becker, S. D., Curran, M. K., Van Horne, A. J., & Fey, M. E. (2015). The efficacy of recasts in language intervention: a systematic review and meta-analysis. *American journal of speech-language pathology*, *24*(2), 237–255. https://doi.org/10.1044/2015_AJSLP-14-0105

Hanen (2011). Early language program. Six steps to follow the child's lead. http://www.hanen.org/Images-for-public-site/Links---Sample-PDFs/TTS_1-ENS.aspx

Girolametto, L., Weitzman, E., Greenberg, J. (2000). Teacher Interaction and Language Rating Scale (TILRS). The Hanan Program. The Hanen Centre: Toronto.

Appendix 5: Descriptive information about children

Table 4 provides descriptive information, based on parent report, on child baseline information as reported by parents.

Results are broken down by group (Early Starter or Late Starter) as well as by LA.

Table 5 Child baseline descriptive data

		Nottinghamshire						**Hackney**					
		Early Start		*Late Start*		*All*		*Early Start*		*Late Start*		*All*	
		N	*%*	*N*	*%*	*N*	*%*	*N*	*%*	*N*	*%*	*N*	*%*
Child gender	*Female*	43	36.8	29	42.6	72	38.9	26	41.3	27	45.0	53	43.1
	Male	74	63.2	39	57.4	113	61.1	37	58.7	33	55.0	70	56.9
Mother's education	*GCSE (age 16) or below*	37	31.6	18	26.5	55	29.7	8	12.5	7	11.9	15	12.2
	A-level (age 18)	20	17.1	14	20.6	34	18.4	12	18.8	11	18.6	23	18.7
	Undergraduate degree	26	22.2	17	25.0	43	23.2	14	21.9	16	27.1	30	24.4
	Postgraduate or higher	21	17.9	17	25.0	38	20.5	23	35.9	22	37.3	45	36.6
	Not sure	13	11.1	< = 5	< = 5	<=20	<=10	7	10.9	3	5.1	10	8.1
Father's education	*GCSE (age 16) or below*	28	24.1	23	33.8	51	27.7	7	11.1	12	20.7	19	15.7
	A-level (age 18)	17	14.7	20	29.4	37	20.1	10	15.9	10	17.2	20	16.5
	Undergraduate degree	26	22.4	11	16.2	37	20.1	20	31.7	14	24.1	34	28.1
	Postgraduate or higher	19	16.4	10	14.7	29	15.8	17	27.0	15	25.9	32	26.4
	Not sure	26	22.4	4	6	30	16.3	9	14.3	7	12.1	16	13.2
Language(s) at home	*English only*	100	85.5	60	88.2	160	86.5	33	51.6	24	40.0	57	46.0
	English, also other language(s)	<=20	<=15	<=10	<=10	<=25	<=15	18	28.1	24	40.0	42	33.9
	Mostly other than English	< = 5	< = 5	< = 5	< = 5	<=5	< = 5	13	20.3	12	20.0	25	20.2
Child age (months)	*Mean (SD)*	30.3(4.1)		29.4(3.1)		30.0(3.8)		31.2(3.9)		31.4(3.8)		31.3(3.8)	
	Range	24.4-38.9		24.1-36.8		24.1-38.9		24.4-38.7		24.2-39.0		24.15-39	
ELIM-E total score	*Mean (SD)*	29.1(14.6)		29.8(13.8)		29.3(14.3)		33.7(14.9)		33.8(14.6)		33.7(14.7)	
	Range	1-50		4-50		1-50		2-50		1-50		1-50	

Appendix 6: Descriptive information about practitioners

Table 5 provides descriptive information, based on practitioner self-report, on baseline information about practitioner background and self-assessed confidence and interactions in support of children's language development.

Results are broken down by group (Early Starter or Late Starter) as well as by LA.

Table 6 Practitioner baseline descriptive data

		Nottinghamshire			**Hackney**		
		Early Starter	*Late Starter*	*All*	*Early Starter*	*Late Starter*	*All*
Highest qualification	Level 2	<=3	<=3	<=3	<=3	<=3	<=3
	Level 3	18	13	31	15	16	31
	Level 4 +	5	<=3	8	10	5	15
Years of experience*	Mean (SD)	8.80(6.87)	6.66(4.92)	7.89(6.14)	8.71(7.42)	9.51(9.62)	9.06(8.37)
	Range	2-25	0.25-17	0.25-25	0-30	1-39	0-39
Total		<=30	<=20	<=50	<=30	<=25	<=50

**1 Early Starter practitioner in Nottinghamshire and 1 Late Starter practitioner in Hackney did not report years of experience at baseline.*