Empowering staff to enhance oral language in the early years: Cluster randomised trial

FINAL REPORT

May 2023

Julie E. Dockrell[1], James Law[2,4], Sandra Mathers[3], Claire Forrest[1], Jenna Charlton[2], Keeley Dobinson[1], Liz Hewitt[3]

[1]Psychology and Human Development, IOE UCL's Faculty of Education and Society.

[2]School of Education, Communication & Language Sciences, Newcastle University.

[3]Department of Education, University of Oxford.

[4] James Law sadly passed away in October 2021

UCL IOE – FACULTY OF EDUCATION AND SOCIETY

Newcastle University

UNIVERSITY OF OXFORD

Nuffield Foundation

Contents

I. Tables and Figures

II. List of Abbreviations

ANOVA	Analysis of Variance
BAS	British Ability Scales
BAU	Business-as-usual
DLL	Dual Language Learner
GAPS	Grammar and Phonology Screening
ICC	Intra-class Correlation
IDACI	Income Deprivation Affecting Children Index
IPE	Implementation and Process Evaluation
LUI	Language Use Inventory
MANCOVA	Multivariate Analysis of Covariance
PD	Professional Development
QUILS	Quick Interactive Language Screener
RA	Research Assistant
RCT	Randomised Controlled Trial
RNAR	Rehearse, Notice, Analyse and Refine
SC	Story Conversations
SEN	Special Educational Needs
SDQ	Strengths and Difficulties Questionnaire
TT	Talking Time
WP	Word Play

III. Acknowledgements

The authors wish to thank the nursery staff, parents/carers and children who participated in the study. We thank Wendy Lee and Clare Williams for their contributions to the development of the manual and professional development; and Wendy Lee and Theresa Redmond for delivering the professional development and supporting intervention schools. We also thank the following Research Assistants who contributed to data collection, scoring and data input: Fernanda Ahumada, Georgia Hellard Timm, Phoebe O'Donnell, Lea Watson, Angelina Vlasenko, Kelly Ng, Annabelle Piot, Alyssa Williams, Jennifer McMahon, Minhee Han, Yasin Arslan, Em Richardson, Louise Chandler, Martha Hallett-Morley, Emily Preston, Rebecca Kemp, Naomi Rose, Naily Khairul-Effendy, Rebecca Posen.

We thank Nikki Hawley for administration support and Maria Cockerill for recruitment in the Northeast. We thank the Steering Group for sharing their expertise and providing feedback on an initial draft of this report.

We are grateful to the following publishers who provided books at a discounted rate which made this research possible: Alanna Max (*Lulu Loves Flowers*); HarperCollins (*Goodnight, Gorilla*); Macmillan (*Peace at Last*); Old Barn Books (*Bee and Me*) and Oxford University Press (*Christopher's Bicycle*). We are also grateful to Brookes Publishing for providing the Quick Interactive Language Screener (QUILS) at a reduced price.

Finally, we are grateful to the Nuffield Foundation for funding this study. The Nuffield Foundation is an endowed charitable trust that aims to improve social wellbeing in the widest sense. It funds research and innovation in education and social policy and also works to build capacity in education, science and social science research. The Nuffield Foundation has funded this project, but the views expressed are those of the authors and not necessarily those of the Foundation. More information is available at www.nuffieldfoundation.org.

Empowering staff to enhance oral language in the early years: Cluster randomised trial

1. Executive summary

Talking Time© is a universal intervention targeting the oral language skills of children aged between 3 to 5 years of age. The intervention is designed to support early years' practitioners to deliver a programme of engaging structured activities to children in small groups in the nursery setting. *Talking Time*© was originally conceived and implemented in a quasi-experimental study in three nursery schools in London (Dockrell et al., 2010). The manual and professional development (PD) were expanded to incorporate recent research evidence. The updated programme was designed to equip staff with the knowledge and skills needed for flexible, high-quality implementation using evidence-based ways of talking with children, and support them in adapting and embedding the programme into regular classroom practice.

After postponement of the first trial due to the COVID-19 pandemic (2019-2020), the current cluster randomised controlled trial (RCT) ran from September 2021 to July 2022. Forty nursery schools from the lowest quintile of deprivation in London and the Northeast were recruited, with a final sample of 36 schools randomised into intervention or business-as-usual (BAU) groups. Although significantly challenged by the ongoing COVID-19 pandemic, the trial was completed with reduced numbers of planned participants (children and settings), alterations to the format of the staff training and with a shortened intervention period. Baseline and follow-up data captured the effect of the intervention on child language, child behaviour and practitioner-child interactions. Analyses evaluated the main effects of Time (Baseline vs Follow-up); Condition (Intervention vs BAU) and Language Status (Monolingual English vs Dual Language Learners). The success of the intervention was determined by a significant interaction between Time and Condition – hereby referred to as the intervention effect - demonstrating larger gains for the intervention group compared to the BAU group at follow-up. The intervention effect was measured on three different outcome levels: assessments of expressive vocabulary, grammar, comprehension and oral narrative skills at the child level; improved adult-child interaction during book reading at the practice level; and improved parent/carer-report of language and behaviour. An implementation and process evaluation (IPE) captured programme implementation, adaptation and future feasibility via staff reflections, surveys and interviews.

Key findings:

- Baseline child data clearly demonstrated the need for a targeted universal intervention to support oral language in the early years. Nursery children in these deprived areas were below average on measures of receptive and expressive language; approximately

17%[1] of parents/carers reported a concern with their child's language; and only 37% of children were read to every day at home.

- Dual language learners (DLLs) had poorer oral language performance than their English-speaking peers but appeared to have strengths in phonological awareness at baseline.
- There was significant growth over time on standardised language tasks for all children (i.e. those in the intervention and BAU groups). Children who were DLL demonstrated greater improvement in receptive and expressive language than their peers. This potentially highlights the benefit of attending nursery for children who are DLL, to help improve their English language skills.
- Over and above children's general language growth, the *Talking Time*© intervention significantly improved children's expressive vocabulary, assessed by a bespoke measure of target object and action words. This is noteworthy because the target words were *implicitly* embedded within the flexible intervention activities and materials: staff were unaware of the target words and their introduction and use was not prescribed; and vocabulary was introduced via shared book reading and guided play rather than being formally taught. There was also a trend for improvement in children's expressive language as measured by a sentence repetition task.
- No effects were identified for other standardised measures of oral language. However, these findings cannot be considered conclusive. The challenges of COVID-19 meant that the intervention programme did not run fully as intended, despite huge efforts and great commitment from participating schools. Recruitment was also affected. Further research under less challenging conditions, with a larger sample size – and also potentially a longer time frame and additional follow-up time-points - are needed to examine effects for the standardised language measures.
- Parent/carer reports of language use identified a significant intervention effect for items identified on the home language scale for children who were DLL. With a small return rate this finding should be interpreted with caution, but it could indicate some transfer effects of the intervention, suggesting that exposure to quality adult-child interactions in early years settings in English has an effect on children's reported language development in their home languages.

[1] All percentages are rounded to the nearest whole number to aid legibility.

- Participating practitioners were asked to share a story book with children at baseline and at the end of the study. In the baseline measure, nursery staff demonstrated a range of practices. An intervention effect was identified for the average number of conversational turns and the proportion of prompts used. Staff in the intervention group also reduced the extent to which they read directly from the text in comparison to the BAU group. These findings indicate the *Talking Time*© intervention influenced a shift in use of oral language exchanges with children and encouraged practitioners to follow the children's lead more.
- Nursery practitioners' perspectives supported these positive findings, with staff reporting benefits of intervention for both practice and child outcomes, particularly for children with additional language needs.
- Overall *Talking Time*© was well-received by schools and was feasible to implement. Nursery staff reported finding both the materials and professional development helpful, with the mentoring element identified as particularly supportive. Staff reported using programme materials flexibly as intended and that their confidence in adapting activities and materials grew over time. Delivery of the programme among the majority of schools was high.
- However, wide-ranging implementation challenges were also highlighted. Levels of COVID-related staff and child absence were high and schools varied in the degree with which they implemented the programme. As noted above, these may have influenced the potential of the programme to impact on child language outcomes.
- While the primary constraints of time, staffing and child absence were related to (or exacerbated by) COVID-19, others related to wider school, staff, child or programme factors. These offer learning for how best to support practitioners in future programmes, and for the future refinement of *Talking Time*©.
- Overall, the Talking Time© intervention produced positive results in challenging circumstances. The Talking Time© programme had a significant positive impact on targeted expressive vocabulary embedded within the activities indicating that language-supporting adult-child interactions can occur through provision of a flexible and an adaptive intervention programme. No intervention effects were identified for the standardised measures of language, although there was a trend for children in the intervention group to have greater gains on the Sentence Repetition task. Importantly adult talk in the intervention groups changed significantly. Intervention practitioners

invited more contributions from children and used techniques for extending the length of these conversations. By corollary there was significantly more talk by the children in these group discussions in the intervention group. Future evaluation in less challenging contexts – and assessing a wider range of practice and child outcomes - is needed to further examine the potential of the approach.

2. Scientific background and rationale

Children from areas of social disadvantage on average perform more poorly on standardised assessments of oral language than their more advantaged peers, setting trajectories which result in poorer performance through school and beyond. The extent and impact of poorer performance has led researchers and practitioners to propose a multi-tiered approach to support children's oral language competence in educational contexts (Ebbels et al., 2019). The first step in a multi-tiered approach involves providing universal evidence-informed support for oral language (see for example Petersen et al., 2022). If children do not progress satisfactorily more targeted effective interventions can be implemented (tier 2) (such as NELI Hulme et al., 2020). A multi-tiered approach has the potential to offer equitable and cost-effective support where children who do not respond to quality first language learning opportunities (see Dockrell. Bakopoulou, Law, Spencer & Lindsay, 2015) are provided with additional targeted support.

In this study we examined the impact of a universal language focussed intervention (*Talking Time©)* designed to support early years practitioners in providing systematic language learning opportunities and interactions to enhance preschool children's language. Using a randomised controlled trial targeting children in the bottom quintile of social disadvantage we examined the impact of *Talking Time©* on children's language performance, on the ways in which staff talked with children, and on the language supporting activities that were provided to the children.

In this section of the report, we outline the rationale for focussing on preschools in areas of social disadvantage; the key language skills that children need and the factors known to support their growth; and barriers to implementing effective language learning pedagogy in educational settings. We then outline how a universal intervention, in this case *Talking Time©*, might address these challenges. Finally, we describe the operationalisation of the programme and the aim and objectives of the current trial.

Differential disadvantage

There has been a long-standing concern about the oral language skills of children from areas of social disadvantage when they enter formal education (at the age of five in England). Low oral language levels on school entry place students at a significant risk of reading and writing difficulties throughout their schooling (Shanahan et al., 2006). Some children are more at risk of poor language development. Universal community surveillance estimates prevalence rates of language delays as between 3% to 8 % of the population at 30 months of age (Sim et al., 2013), with children from areas of social disadvantage experiencing disproportionate delays

relative to their more advantaged peers (Law et al., 2018; Nelson et al., 2011). Preschool oral language skills at kindergarten entry explain most of the effect of social disadvantage on elementary school performance (Durham et al., 2007). Environmental opportunities in homes and schools are the main constraining factor in the language development of children from areas of social disadvantage (Dietrichson et al., 2017).

Children whose families speak a language other than the language of instruction (English in the case of this project) at home can be at risk of not developing proficient language skills. Children whose language is not the official language of instruction are a heterogeneous population, including migrants, refugees, and children born in the country but whose home language is different from the language of instruction. For many children, language status is highly correlated with socio-economic status, with individuals not born in their country of residence being twice as likely to suffer deprivation as native residents (Eurostat, 2018). Furthermore, adults with low educational attainments (overrepresented within the immigrant population) are three times more likely to be at risk of poverty than those with tertiary education (Eurostat, 2019). Children growing up in these contexts suffer a double jeopardy of low language status and poverty (OECD, 2018). For preschool children where English is an additional language, both the amount and the nature of children's exposure to the language of instruction is a central driver of subsequent language and literacy skills (Buysse et al., 2014).

Language delays have also been associated with a range of social difficulties including peer problems and hyperactivity. These difficulties have long term impacts on later opportunities and mental health (Law et al., 2009). The few studies that have examined the link between oral language and social difficulties in socially disadvantaged areas have found similar significant associations between poor oral language and higher rates of behavioural problems in preschool and primary children (Huaqing Qi & Kaiser, 2003; Law et al., 2014; Sim et al., 2013). Law and colleagues found a mediating effect of pragmatic language on the association between core language ability and social, emotional and behavioural difficulties (SEBD) in monolingual English-speaking children between the ages of five and 12 (Law et al., 2014).

Consistent evidence of the importance of early oral language skills has led to a concentrated effort to enhance young children's competence in both expressive and receptive language. While there is widespread agreement that additional support is needed to address these inequities, there is continued disagreement about the ways in which this should be done. The use of child health surveillance programmes to capture language delays (Law, Charlton,

& Asmussen, 2017), targeted parental packages (Davies et al., 2020) and specific language interventions (Fricke et al., 2013) have all been proposed as ways of addressing the impact of children's language delays (see also Ebbels et al., 2019). These targeted approaches have had varying degrees of success but overall point to the importance of ensuring that all children are provided with evidenced-informed universal language support when they enter education (see for example Dobinson & Dockrell, 2021). Given the disproportionate numbers of children in at risk contexts with poor oral language skills there is a need to provide more systematic "targeted" universal interventions in educational contexts (see Law et al., 2017; Law, 2019). Despite this reported need and a focus on multi-tiered approaches, limited use of language supporting strategies have been recorded in settings in areas of social disadvantage (see for example Phillips et al., 2018). Education professionals need to know which language skills should be the focus of their attention and how language development can be supported.

Key components of oral language in preschool children: what should be focused on

An understanding of the key components of oral language is necessary to support children who struggle with oral language and to provide effective language learning environments to support literacy and learning, (Oracy All-Party Parliamentary Group (APPG), 2021). Growth in oral language during early childhood reflects a continuous development of lexical representations (vocabulary) and the development of an implicit understanding of the rules of grammar; skills which have been described as core language components (Karlsen et al., 2021). In addition to these core language components the ability to create a narrative and engage in connected discourse has been identified as an important precursor for social communication and literacy skills (Griffin et al., 2004). All three language components are reported to be compromised in many children from areas of social disadvantage.

Most research studies have focused on vocabulary, leading to an emphasis on *'the word gap'* (Golinkoff et al., 2019). Children from low-income households often begin school with vocabulary levels significantly below that of their more advantaged peers (Hoff, 2003; Huttenlocher et al., 2010). Developing an extensive receptive and expressive vocabulary (vocabulary breadth) is essential for language and literacy development. It is of equal importance to develop knowledge of the lexical/semantic relationships between words so that conceptual development is enhanced (Wojcik, 2018). As children's vocabulary increases, they learn the ways in which words are related to each other (vocabulary depth). These early lexical semantic relationships are organised both categorically and thematically, and embedding this distinction in vocabulary interventions is important (Neuman & Dwyer, 2011). Vocabulary breadth and depth are related to different features of reading, with

vocabulary breadth predicting reading decoding and vocabulary depth predicting reading comprehension (Ouellette, 2006). Thus, developing both the depth and breadth of children's vocabulary should be embedded in educational practice.

Vocabulary is only one dimension of the language system that is, on average, challenged for children from areas of disadvantage (Levine et al., 2020; Weiler & Decker, 2022). Given the conceptual and empirical relationship between children's vocabulary knowledge and grammatical competency (Language and Reading Research Consortium, 2015; 2017; Karlsen et al., 2021; Tomblin & Zhang, 2006) it is not surprising that comparable delays in grammar are also evident (Levine et al., 2020). Arguably the child requires a critical mass of words for grammar to emerge. Vocabulary knowledge provides a foundation for learning and producing syntactic structures. Of particular importance to this next step in language development is the use of inflectional morphology; that is information included in a word to reflect tense or plurality. Hirsch-Pasek and colleagues identified what they called a 'grammar gap' (Hirsh-Pasek et al., 2015, as cited in, Pace et al., 2017). In their study at five years of age children from areas of social disadvantage had, on average, the grammatical competency of a 24-month-old; that is some three years behind their more advantaged peers. There is also growing evidence that children from areas of social disadvantage may struggle in acquiring new linguistic concepts as a result of poorer language learning process skills (Levine et al., 2020; Weiler & Decker, 2022), which likely impact on developmental trajectories of core language skills and the types of support that will be needed in educational settings.

Vocabulary and grammar underpin the development of connected discourse, such as conversations or narratives. Discourse skills develop as children engage in interactions with adults who provide scaffolding and feedback to develop these skills. Research has shown that difficulties in producing extended discourse and using language to engage others limits children's ability to communicate with others in social settings and to actively engage with classroom activities (Apel & Apel, 2011; Snow, 2014). Not surprisingly, given the challenges with vocabulary and grammar for many children from areas of social disadvantage, narrative skills are also compromised.

The language learning context is key to supporting oracy skills in the early years. Research has highlighted the lack of training around oral language skills in practitioners (Dockrell et al., 2017) and the variable quality of the experiences children receive to support the development of their oral language skills in preschool settings (Dockrell et al., 2015).

Understanding the factors which support the development of children's language skills becomes an important driver of reducing the attainment gap.

Supporting language development and barriers to impact

The linguistic environment that children experience and interact with enhances their developmental trajectories (Hoff, 2006). Evidence suggests that language growth is optimised by adult input which incorporates interactive, linguistic and conceptual elements, in particular: talking *with* rather than *to* children; increased diversity and complexity of oral language during early childhood; and a gradual transition from contextualised to decontextualised conversations (Rowe, 2022; Rowe & Snow, 2020). These factors are malleable and can be embedded within early years settings. For example, educators' use of sophisticated vocabulary, expansions of children's language and use of literal and inferential questions all predict child language growth (Cleave et al., 2015; Dickinson & Porche, 2011; van Kleeck et al., 2006), with social contingency a key feature (Cabell et al., 2015; Duncan et al., 2020). Multi-turn conversations are understood to offer rich contexts for language growth, in part because they promote joint attention and increase opportunities for adult's contingent use of language within a socially meaningful context (Girolametto & Weitzman, 2002; Romeo et al., 2018).

Certain classroom contexts offer unique potential for language learning. For example, interactive book reading offers opportunity for supporting word learning and extended conversations (Girolametto & Weitzman, 2002; Zucker et al., 2013) and has been shown to improve receptive and expressive vocabulary and narrative skills (Cabell et al., 2019; Grolig et al., 2020; Mol et al., 2009; Pesco & Gagné, 2017). Small group activities increase opportunities for joint attention (Hassinger-Das et al., 2017), encourage children to interact and make verbal contributions (Pellegrino et al., 1990) and encourage educators to offer higher-quality language-supporting practice matched to children's developmental levels (Turnbull et al., 2009). Guided play following shared book reading, and approaches which combine implicit and explicit vocabulary instruction, have also been shown to enhance receptive and expressive vocabulary (Marulis & Neuman, 2010; Toub et al., 2018).

As noted, in areas where high proportions of children are at risk of oral language delay there is a need for a 'targeted universal' approach, with language-learning interactions and opportunities offered to *all* children as part of the regular classroom day. However, observational data indicate that language-supporting practice in English schools serving disadvantaged populations may be insufficient to meet children's language needs (Dockrell et al., 2015; Law et al., 2019). Areas of weakness include the lack of structured small groups,

interactive book reading, limited language supporting interactions and reduced encouragement of children to talk with others (Dobinson et al, 2022; Wright et al, 2020).

Evidence also suggests that the professional development which might strengthen practice is inconsistent in its quality (Cordingley et al., 2015) and in its impact on teaching quality or child outcomes (Markussen-Brown et al., 2017); and that educators experience a lack of training around oral language (Dockrell et al., 2017). There is a clear need for interventions which are informed by evidence on professional learning as well as child development and pedagogy. Moreover, there is a need to move beyond a focus on vocabulary and reducing the *'word gap'* in interventions (Hindman et al., 2016), and to offer opportunities for extended discourse within early years classrooms (Law et al., 2019).

Why Talking Time?

Talking Time© is a universal intervention targeting children's vocabulary, narrative and comprehension skills between the ages of 3 and 5. It aims to support early educators in offering engaging activities and interactions designed to provide optimal conditions for oral language growth. Specifically, it offers meaningful and developmentally appropriate structured small-group experiences which allow children to hear and use language in the context of socially contingent multi-turn conversations. These include:

- interactive book reading using books (wordless and text-and-picture) selected to promote vocabulary, grammatical complexity and conversation.
- engaging vocabulary activities which combine explicit and implicit instruction; and
- construction and retelling of narratives based on familiar situations.

These activities were designed to create opportunities for talk which is meaningful to children and child led. They were reinforced by follow-up classroom activities in order to promote deep and sustained language growth across multiple contexts (Harris et al., 2011).

Activity context naturally shapes adult input. For example, educators use more complex language during reading (Hoff, 2010) but more contingent responses during play (Røe-Indregård et al., 2022). Wordless books elicit greater educator instructional support and child language input but text-and-picture books elicit longer sentences from educators and support lexical and grammatical diversity (Chaparro-Moreno et al., 2017). Prior shared reading and playful vocabulary interventions have demonstrated benefits for child language (Dowdall et al., 2020; Marulis & Neuman, 2010). However, moderator analyses indicate greater effects when adults are *also* trained to use specific pedagogical strategies (Marulis & Neuman, 2010). The original manualised version of *Talking Time*© (Dockrell et al., 2010) was thus extended for this trial to include a more comprehensive professional development

(PD) component. The updated programme was designed to equip staff with the knowledge and skills needed for high-quality implementation and support them in adapting and embedding the programme into regular classroom practice. The rationale for the PD design is presented below.

The evidence-base underpinning the Talking Time© professional development (PD)

Effective PD provides educators with the domain-specific pedagogical knowledge which underpins effective practice (Coe et al., 2014; Desimone, 2009; Shulman, 1987). Such knowledge is operationalised in *Talking Time©* via a framework of eight language-supporting strategies shown to facilitate oral language growth (as described above) which underpins the professional development (see Appendix A3).

Effective PD also supports application in practice (Zaslow et al., 2010). The *Talking Time©* manual provided flexible activity prompts and plans to support educators in implementing programme activities using the language-supporting strategies. This was based on evidence that prompts and cues help to scaffold and embed new behaviours (Sims et al., 2021) and evaluations of manualised early childhood curricula which showed impacts for child language growth (Weiland & Yoshikawa, 2013). However, *Talking Time©* also aimed to nurture the explicit procedural knowledge and skills which educators need to adapt the programme and to embed it into routine practice in the longer term (Anderson & Krathwohl, 2001; König, 2013). By combining initial training with instructional coaching (both in-person and via video) it offered explicit instruction in the language-supporting strategies alongside opportunities for rehearsal in practice. Instructional coaching (Joyce & Showers, 1981; Kraft et al., 2018) is understood to be one of the most powerful PD approaches because it reflects many evidence-based mechanisms for developing teaching techniques. These include modelling; observing and providing feedback; opportunities to rehearse techniques in practice; supported reflection; and motivating goal-directed behaviour through goal-setting and positive reinforcement (Elek & Page, 2019; Sims et al., 2021). Feedback and analysis using video interactions also has been shown to support growth in educators' pedagogical knowledge, skills and practice (Fukkink & Tavecchio, 2010; Sherin & Van Es, 2009).

Throughout the PD, the language strategy framework was used as a reflective tool to support educators in rehearsing, noticing, analysing and refining their practice (Figure 1). This draws on evidence that context-specific repetition supports the embedding of new practices (Sims et al., 2021) and that educators' ability to notice and analyse practice predicts child language growth (Matherset al.,2022). The use of structured rubrics to observe and

analyse practice has been shown to lead to improvements in both practice and child outcomes (Burgess et al., 2021; Wright et al., 2020; Pianta et al., 2017; Reddy et al., 2021).

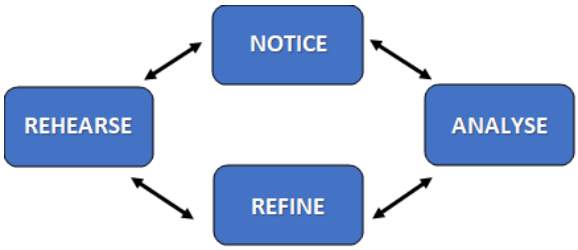

Figure 1: The Rehearse, Notice, Analyse and Refine (RNAR) cycle used to embed language-supporting practice

Prior studies show that the vast majority of interventions are not implemented with fidelity and are often adapted by educators in ways which are inconsistent with the intended design (Piasta et al., 2015). *Talking Time*© was designed to support 'principled adaptation' (Durlak & DuPre, 2008; Lendrum & Humphrey, 2012) by offering explicit guidance for adaptation in alignment with key programme principles (Sims et al., 2021). This combination of flexible curricular support alongside support for professional learning is reinforced by studies showing that lightly scripted programmes incorporating PD promote greater language growth than fully scripted programmes, likely because they allow educators to tailor activities for individual needs (Bleses, Højen, Dale, et al., 2018). The contained professional learning context provided by *Talking Time*©; and repeated use of the RNAR cycle to embed the framework of language-supporting strategies (Figure 1) was theorised to provide a highly efficient context for learning and behaviour change. In essence, the intervention was used as a vehicle for developing longer-term knowledge and skills which support child language growth in a manner tailored to individual needs. Finally, the programme draws on learning from implementation science (e.g. Bleses et al., 2021; Durlak & DuPre, 2008) in scaffolding effective and sustained delivery; for example, by explicitly supporting school readiness for the programme, alignment with school practices and goals, and longer-term leadership and sustainment. The operationalization of the *Talking Time*© programme is outlined in section 3.5.

2.1 Aim and objectives

The *Talking Time*© intervention aims to support language learning for all children in the classroom by embedding the activities within every-day practice and supporting staff to develop ways of talking with children. The current trial evaluated the efficacy of the intervention in the following ways:

1. We expected the primary outcomes of child structural language (expressive vocabulary, grammar, comprehension and oral narrative skills) to significantly improve in the intervention group. We did not expect any intervention effects to be identified for measures included as control variables (phonological awareness and non-verbal ability).
2. We anticipated that the *Talking Time*© intervention would have a significant impact on the practice of participating nursery staff, changing the ways they engage and interact with children.
3. We predicted that parent/carer report of language use and behaviour would change over time, but we did not expect these secondary outcomes to be impacted by the *Talking Time*© intervention.

3. Method

3.1 Ethics

Ethical approval was granted by the IOE, UCL's Faculty of Education and Society Research Ethics Committee (REC 1118: Empowering Staff to Enhance Oral Language in the Early Years).

3.2 Trial registration

3.2.1 Original trial

The original trial was registered 20.09.2019 on the American Economic Association's registry for randomised controlled trials. RCT ID: AEARCTR-0004379.

3.2.2 Amendment

The trial registration was updated 22.10.2021 to reflect the new start date (1.7.2021) and end date (31.12.2022) after the funding was granted again following the impact of COVID-19 on access to schools and children. The baseline date, intervention date and follow-up date were also revised. Amendments were also made to reflect the adaptations made to the method at child level and school level as mentioned above in section 2.1. The amended trial registration can be found online at https://doi.org/10.1257/rct.4379-2.0.

3.3 Design

The study took the form of a two-armed, stratified (North South) cluster randomised controlled trial. Clusters were at the school level. Schools were randomly allocated to either the intervention arm (received training in *Talking Time*© in spring, 2022) or the business-as-usual (BAU) arm (waitlist until after study conclusion in autumn, 2022). Business-as-usual was chosen as the comparator given the universal nature of the intervention and the aim to compare with typical early years practice in nursery classes. Temporary Research Assistants

responsible for collecting child level data at baseline and follow-up were blind to the allocation of schools.

3.4 Schools' eligibility criteria

Maintained primary and nursery schools in the lowest quintile of deprivation in London and the Northeast of England were recruited via flyers and presentations to school partnerships and early years leaders. Schools were eligible for inclusion if they were within the lowest quintile of deprivation. Deprivation information was generated from school postcodes using the Income Deprivation Affecting Children Index (IDACI) from the English Indices of Deprivation 2015 (Smith et al., 2015). The IDACI rank ranges from 1 (most deprived) to 32,844 (least deprived). Schools exceeding this cut-off, or which had recently introduced a language intervention not considered part of usual practice in the nursery class, were not eligible for the study. Twenty-eight schools in London expressed interest, of which 21 met inclusion criteria for deprivation. One was excluded because they had recently implemented a language programme in the nursery. Ninety-seven schools meeting inclusion criteria were approached in the Northeast. Seventy-two did not respond to invitations or declined to participate. Three schools were not eligible to participate due to recently implementing another language programme and two declined to take part due to limited staffing capacity. In total, 20 schools from London and 20 from the Northeast were recruited into the study at baseline. Within each school, the nursery class (providing for children aged 3 to 4 years of age) participated in the study.

3.5 Intervention

Programme activities, materials and schedule

The *Talking Time*© intervention programme comprised three small-group activities:

- **Story Conversations (SC):** adult-child co-construction of stories and conversations using the book illustrations as prompts (a form of interactive book reading).
- **Word Play (WP):** engaging activities designed to rehearse and reinforce vocabulary and concepts introduced during SC. Sessions were a mix of guided role play and targeted games designed to reinforce specific words, word types or concepts.
- **Hexagons:** narrative discussion and retelling based on photos of real situations likely to be familiar to children (e.g., going shopping, going to the doctor).

The aim was for all children to take part in two 15-minute activities per week during regular class time, in mixed language-ability groups of five children or fewer. The programme lasted 16 weeks (timetable shown in Appendix A1) with children participating in 15 SC, eight WP and five Hexagons sessions during this period. A longer programme had

been intended but was shortened due to COVID-19-related delays (see Appendix A). Schools were also encouraged to provide at least one reinforcement activity within their regular classroom practice each week.

SC and WP sessions were based on five picture books (some wordless, some text-and-picture) with three weeks spent on each book (see Appendix A2). Schools were also provided with a starter pack of seven pre-prepared Hexagons photosets and resources to create their own sets to represent the locality and community activities.

Implementation support system: prompts and plans

The *Talking Time*© manual provided a resource bank to support activity planning and implementation. For each SC book and Hexagons photoset, possible target vocabulary, concepts, conversation topics, narrative elements (Hexagons only) and props were suggested. A range of conversation prompts was also provided for each book page/photo. These were designed to elicit the language-supporting strategies represented in the Talking Time framework of language-supporting strategies (see Appendix A3, A4, A5). For example, prompts were provided at different levels of challenge (low, mid, high). Low-challenge prompts included closed questions, completion prompts and forced alternatives and were largely literal. Higher-challenge prompts included open-ended questions and were designed to elicit decontextualised language. Prompts also supported the modelling and rehearsal of vocabulary and narrative skills.

Six possible WP plans were provided for each book: three imaginative role play plans and three targeted games. Across the three weeks spent on each book, schools were advised to use role play for the first two weeks and a targeted game for the third, with plans selected and adapted based on child needs and interests. Plans provided guidance on how to introduce and reinforce key words and concepts and play the game whilst modelling and encouraging children to use vocabulary. Ideas were provided for classroom reinforcement activities.

The manual also provided guidance on how to prepare for activities and made the planning principles explicit to support longer-term self-sufficiency. Staff were encouraged to select and adapt plans, props, vocabulary and prompts in advance as appropriate for each group and context. However, they were guided and supported to use plans *flexibly* during sessions, adapting in response to the children and following their conversational lead (e.g., co-constructing a story *with* children rather than telling the story as shown in the book).

Implementation support system: professional development (PD)

Implementation was further supported by the PD component. Each school was supported by a trainer-mentor (one per area). Both had a speech and language background, and expertise in

supporting practice in early education contexts. The PD elements are shown in Table 1 and timings in Appendix A1.

Table 1: Professional learning components of Talking Time (TT)©.

Aspect	*Attendance*	*Format*	*Quantity*	*Aims + evidence-based PD mechanism**
Twilight training	All staff regularly delivering sessions	Online	3 x 1.5 hrs	Designed to introduce the activity and underpinning pedagogical principles. ***Mechanisms*:** instruction in teaching techniques, modelling.
In-class mentoring	*Talking Time (TT)©* leads	In person	3 x 2 hrs (1 hr per lead)	Designed to support application and embedding. ***Mechanisms:*** instruction, modelling, observation & feedback, context-specific rehearsal, goal setting, positive reinforcement, action planning.
Video mentoring	One *TT©* lead (pedagog-ical leader)	Online	3 x 45m-1 hr	Designed to support application and embedding & longer-term leadership, planning and adaptation. ***Mechanisms****:* as above.
Individual reflection	*TT©* leads	Indiv-idual	Weekly	Designed to support application and embedding. ***Mechanisms:*** context-specific rehearsal, action planning, self-monitoring
Team collabor-ation	As needed	Team	Weekly	Designed to support embedding and sustainment. ***Mechanisms***: social support, developing a professional learning community, context-specific rehearsal, action planning.

*Drawn from literature cited in Section 2 e.g., Sims et al, 2021, Elek & Page, 2019.

The PD was underpinned by a framework of evidence-based language-supporting strategies (the Rehearse, Notice, Analyse and Refine - RNAR - framework) (see section 2 and Appendix A3). During training this was used to analyse exemplar videos of *Talking Time©* groups. During in-class mentoring, staff and mentor observed each other leading activities and used the framework to support professional dialogue, feedback, analysis and goal setting. During video mentoring, mentor and mentee used the framework to analyse a recorded *Talking Time©* session. Educators were also encouraged to complete one individual reflection each week (recorded in their 'Individual Log') and to engage in weekly *Talking Time*-related professional collaboration; for example, shared planning, professional discussion or observation and analysis of sessions using the RNAR framework.

The PD explicitly supported staff in understanding the principles behind the programme, to enable them to adapt plans for individual child needs and gradually decrease reliance on programme materials and support by the end of the structured intervention period. The final section of the manual and the third video mentoring session focused explicitly on

planning, adapting and embedding activities following the end of the structured intervention period, including developing new activities based on materials of the schools' own choice. Mentors also provided support for adaptation throughout.

Supporting successful delivery

A number of wider programme features aimed to support successful implementation, drawing on the literature cited in Section 2. Orientation visits from trainer-mentors on entry were designed to build relationships, clarify expectations, and anticipate and problem-solve practical barriers. In order to ensure clarity of roles and leadership, schools were asked to identify two *Talking Time*© leads (usually a teacher and teaching assistant) to lead on implementation and deliver the majority of sessions. In some schools, additional staff also delivered sessions. Video mentoring was designed to further support the pedagogical leader (usually the class teacher) to lead the wider team in planning and adaptation. Weekly professional collaboration was encouraged (described above) and *Talking Time*© leads were encouraged to cascade learning from in-class and video mentoring. In order to support adherence, staff were encouraged to create a timetable and keep a session log to monitor delivery and child exposure. Finally, in weeks with no training or mentoring, mentors sent weekly emails to maintain contact and momentum.

Theory of change

Key programme components included the *Talking Time*© activities (SC, WP, Hexagons), the manual (activity plans and prompts) and professional development (RNAR framework and support to apply and embed). These were anticipated to influence the primary outcome of children's structural language ability (expressive vocabulary, grammar, comprehension and oral narrative skills) via:

1. Children's exposure to *Talking Time*© activities.
2. The ways staff talked <u>with</u> children and improvements in their knowledge, skills and language-supporting practice.
3. Successful programme implementation in a manner which is flexible but congruent with programme design and which is sustained over time.

3.6 Outcomes

3.6.1 Child measures

At the child level, direct assessments measured the primary outcome of structural language ability before and after the intervention. Assessments of nonverbal ability and phonology were also included as control variables. Adaptations to the outcome measures from the

original trial are discussed in Appendix A. Details of the reliability and validity of tasks are found in Appendix B.

3.6.1.1 Control variables

Phonology. The Non-word Repetition subtest from the Grammar and Phonology Screening test (GAPS; Gardner et al., 2006) required children to repeat eight non-words which vary in complexity (e.g. *dremp, bademper, difimp,* etc.).

Nonverbal abilities. Two subtests from the British Abilities Scale 3rd edition (BAS; Elliot & Smith, 2011) were administered. The Matrices scale required children to choose the missing picture that fits the pattern shown in a 4x4 matrix. The Picture Similarities subtest required children to point to one of four pictures that matched the concept represented in the target picture.

3.6.1.2 Target variables

Grammar. Children repeated eleven sentences that targeted various aspects of syntax in relation to pictures from the Sentence Repetition component of the GAPS (Gardner et al., 2006).

Comprehension. The Verbal Comprehension subtest from the BAS-3 (Elliot & Smith, 2011) was administered as a measure of receptive language. Children pointed to pictures and manipulated objects in response to questions and instructions.

The Quick Interactive Language Screener (QUILS; Golinkoff et al., 2017) was also administered as a measure of receptive language. The QUILS is designed to be administered on a touch-screen device enabling children (aged 3 years to 5 years 11 months) to point to the picture of the correct answer. There are 48 items covering three areas of language: Vocabulary, Syntax and Process (a measure of how well children can learn new information and incorporate it with their existing knowledge). Raw scores, standard scores and percentiles are automatically generated with a maximum possible total raw score of 16 for each area.

Expressive vocabulary. The Naming Vocabulary subtest from the BAS-3 (Elliot & Smith, 2011) required children to name pictures of objects.

To evaluate the effect of the *Talking Time*© programme in more detail a bespoke list of vocabulary items was constructed from the Object and Action Naming Battery (Druks & Masterson, 2000). Eight objects (nouns) and eight actions (verbs) with an age of acquisition of at least three years were selected. Four of each were target words that were featured in the story books used in the intervention. Some of these were represented in the text, while others (e.g., in the wordless story books) were only represented in the illustrations. The eight target words were also included in the Story Conversations and Word Play prompts and plans provided in the *Talking Time*© manual for the relevant book and (where relevant) in the

Hexagons prompts. However, these were not made explicit to practitioners, who had the choice of which vocabulary to focus on and which plans and prompts to use. None of the control words were prominent within the intervention books. When completing the assessment, children were asked to name individual black and white line drawings for each of the words.

Oral narrative skills. At follow-up children listened to the researcher read aloud a story about a "naughty bus", then were asked to retell the story using the pictures as guides (The Bus Story; Renfrew, 1997). The Bus Story was administered on a tablet and children's responses were recorded using an audio recorder, as well as by hand. The Bus Story has three components: Information (the amount of information the child conveys when telling the story), Sentence Length (the mean sentence length of the first five sentences) and Complexity (use of composite sentences). Participants received a score of 2, 1 or 0 depending on the amount of detail that was reported, with a maximum possible total of 54 points.

3.6.2 Practice measures

Practitioners were asked to select a group of four children with varying language competence and complete three activities: sharing a text and picture book ('Storm Whale Reading'); playing with a set of toys chosen to align with the Storm Whale book ('Storm Whale Toy Play') and sharing a wordless book ('Snowstorm Reading'). The two books were: The Storm Whale by Benji Davies and Once Upon a Snowstorm by Richard Johnson. Practitioners were instructed to interact with children as they typically would and to record their interactions using a handheld voice recorder. For each activity practitioners reported group level information (e.g., the number of children who were DLL, the number of children with special educational needs). Practitioners uploaded their recordings and information forms to the secure server online.

3.6.3 Parent/carer reported measures

Secondary outcomes of language use and behaviour were measured via parent/carer report. This allowed for exploration of the children's abilities in their home language and further examination of the association between language and socioemotional development as a secondary outcome. However, as the *Talking Time*© programme was delivered in the classroom setting and did not directly target these variables we only predicted small effects.

3.6.3.1 Parent/carer report of child language ability

The Language Use Inventory (LUI; O'Neill, 2009) is a standardised parent/carer-report measure of language use in daily life for children aged 18–47 months designed to assess

young children's spoken pragmatic language use. For the purposes of the trial a modified version of the LUI was used which shortened the completion time for parents/carers, and prioritised subscales relating to expressive vocabulary and grammar (Dockrell et al., 2022). Additionally, parents/carers were first asked to record their child's abilities in English and then in their home language. Six of the 10 original LUI subscales were included in the adapted version. Three subscales focused on vocabulary (C: Types of words your child uses, F: How your child uses words to get you to notice something, I: Your child's use of words in activities with others) and three focused on more extended language use and grammar (D: Your child's requests for help, H: Your child's questions and comments about themselves or other people, N: How your child is building longer sentences and stories). Parents/carers were asked to complete the LUI-6 by reporting on their children's current language and communication skills using a nominal scale (Yes/No) or ordinal scale (Never/Rarely/Sometimes/Often). Scores of 1 were attributed to items reported as 'Yes' and 'Sometimes' or 'Often'.

3.6.3.2 Parent/carer report of child socioemotional behaviours

The Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997) consists of 25 items that are rated on a scale of *Not true*, *Somewhat True* and *Certainly True*. Scores of 0-2 are assigned to each rating and positive items are reverse scored. Five subscales are produced: Peer Problems, Emotional Problems, Hyperactivity, Conduct Problems and Prosocial behaviour. Each subscale provides a score between 0-10. The first four subscales are summed to create a Total Difficulties score ranging from 0-40. Based on normative scores the degree of difficulty for each subscale and Total Difficulties scale can be categorised as follows: 0-12 'Close to average'; 13-15 'Slightly raised; 16-18 'High' and 19-40 'Very high'.

3.7 Implementation and Process Evaluation (IPE)

The IPE aimed to establish:

1. How *Talking Time*© was implemented; and how and why implementation varied.
2. How and why the *Talking Time*© programme was adapted by schools and to what extent adaptations were congruent with programme principles.
3. Whether and how schools intended to continue with the programme following the end of the structured intervention period and what challenges they may face.

The IPE was based on survey responses, interview data and records from the 14 intervention schools which completed the programme (the 'completers'), one school which dropped out

prior to the programme end (a 'leaver'), nine schools from the business-as-usual (BAU) control group and the two trainer-mentors.

Two mid-programme surveys (Weeks 6, 12) and a final survey were distributed to participating practitioners in intervention schools. All included a mix of Likert scales and open response questions. Mid-programme responses were collected from approximately two-thirds of schools (Week 6 *n* = 12, Week 12 *n* = 13; *n* =20 staff at each point). Final surveys were collected from the 14 'completer' schools (*n* = 25 staff).

Interviews with 14 intervention schools (thirteen completers and one leaver) were completed by a researcher who was not involved in developing the PD. Interviews were conducted online following the end of the intervention period and lasted approximately one hour. Most were conducted jointly with both *Talking Time*© leads. The generic interview schedule was tailored for each school based on survey responses and weekly log data. A separate schedule was used for the 'leaver' school. Short interviews were also conducted with BAU schools (*n* = 9). The two trainer-mentors were jointly interviewed in person to capture their perspectives on the programme and its implementation.

Weekly records of session delivery and attendance were gathered from the 14 completer schools. Ongoing pressures on schools during the intervention period led to gaps in some records (*n* = 9 schools), despite multiple efforts to gather missing data. No records were received from leaver schools (*n* = 4 schools). Mentors were asked to complete records of PD delivery and checklists of content coverage.

Details of all IPE materials can be found in the Implementation Annex.

3.8 Sample size calculations

For the 2019-20 trial a power analysis using G*power to detect an effect size $d > .20$ with an alpha set at $p < .05$ indicated that N = 600 child participants were required for sufficient power (*n* = 300 intervention, *n* = 300 BAU, divided equally between the two sites of London and the Northeast). We anticipated assessment of 20 children per class and, given the common occurrence of attrition in large RCT studies (Bleses, Højen, Justice, et al., 2018), we over-recruited nurseries by approximately 10% with a goal of 36 schools (18 of which would receive the intervention). In the 2019-20 trial, 875 children were recruited from 39 schools with 875 children completing the assessment at baseline (2% attrition). In the retrial we fell just short of the target with 584 children recruited from 38 schools. Despite initially recruiting 40 schools into the study, one school was excluded due to recently implementing a different language programme and one school withdrew due to limited staff capacity before the baseline data collection began. A further two schools withdrew resulting in a final sample

of 36 schools to be randomised. Reduced staff capacity due to the ongoing COVID-19 pandemic and smaller class sizes contributed to the smaller sample size in the retrial (see Appendix F1 for details of class sizes). More details about the constraints COVID-19 placed on the evaluation are included in Appendix A.

3.9 Randomisation

Randomisation was conducted by an independent statistician and followed the treatment allocation by minimisation process (Altman & Bland, 2005), based on a median split for level of deprivation (IDACI z-score) and class size (mean maximum number of pupils in the morning session) in the two areas (London and the Northeast). The process of randomisation had two stages; the first stage assigned a number to each school in the consenting group of 36 schools, between 1 and 36 (i.e., as if the schools were consenting one at a time). The schools were then entered in that number order to the second stage: randomisation to two groups, intervention and BAU, based on the pseudo random minimisation method (Altman & Bland, 2005). The minimisation attempted to balance the groups on three factors: area (two levels – London and Northeast); IDACI (two levels – above and below the median split, Median = -1.34776), and number of children (including all-day children) currently attending the morning session (two levels - above and below the median split, Median = 26). The first school to enter the randomisation did so based on random number draw (above or below 0.5) to one of the groups. Subsequent schools were assigned to the group which would minimise the difference between the groups, with a probability of .75. Any ties were resolved by random number draw, similar to first entry.

3.10 Timeline

Table 2 shows the timeline of the trial. Schools were contacted and assessed for eligibility in spring/summer 2021. Upon signing the memorandum of understanding (MOU) head teachers consented to their school being randomly allocated to either the intervention or BAU group and committed to supporting staff participating in the trial (see Appendix C for MOU and consent). Research Assistants (RAs) were recruited and trained in September (see Appendix D for details) and baseline child assessments began in the first week of October 2021 and were scheduled to conclude before the end of term in December 2021. Due to child and staff absence reflecting the ongoing impact of COVID-19, some baseline child assessments continued into January 2022 but were completed prior to the start of the intervention.

At the school level, practitioners recorded their consent and demographic information on a secure server (REDCap) in autumn 2021. Each school nominated two *Talking Time*©

leads who would receive *Talking Time*© mentoring support and deliver the majority of the programme should their school be allocated to the intervention arm of the trial. Parents/carers were given the option for their child to opt out of the audio recording.

Randomisation was conducted by an independent statistician in December 2021 and care taken not to divulge the allocation of groups to the Research Associates (CF & JC) until all the baseline data had been collected. Research Assistants who collected the data remained 'blind' to groups throughout the trial. The intervention ran until June 2022. Follow-up data collection of child, parent/carer and practice measures ran between May and July 2022, following the same procedure as baseline. The time between baseline and follow-up assessments was approximately seven months.

Table 2: Timeline of the trial.

Time	Activity
Spring/Summer 2021	Recruitment of schools
September - October 2021	Nursery staff invited to a PD session (1.5 hours) about language development (online or face-to-face) to familiarise all schools with the importance of oral language and ways to support oracy Consent forms delivered to parents/carers via nursery staff Consent forms collected and children assigned unique ID Recruitment and training of Research Assistants
October – January 2022	Baseline child assessments administered Baseline data entry Audio recording activities to assess small group practice
December 2021	Schools randomly allocated to intervention or BAU
December 2021 – January 2022	Intervention schools met with trainer, received resources
January 2022	BAU schools received story books
January – June 2022	Intervention
May - July 2022	Follow-up data collection

3.11 Analysis plan

3.11.1 Child Measures

An intention to treat analysis was used to examine all directly assessed child measures.

Despite COVID-19 reducing in Autumn 2021 the intervention was impacted by recruitment of schools, smaller nursery class sizes and fewer parents agreeing to participate, as compared with the earlier aborted (2019-20) trial. The maximum number of allocated places in each morning class ranged from 9 to 45 (M = 28.22, SD = 7.95). In 16 schools (n = 6 London; n = 10 Northeast) less than half of the class consented into the study. Thirteen of the schools had more than 50% of the class consent and 7 of the schools had 80% or more consented into the study (see Appendix F1 for details). As a result, we under recruited for power for some comparisons and have insufficient sample sizes to analyse across schools. To address these limitations, we assessed the impact of the intervention using repeated measures analyses which allowed us to capture both change over time (baseline vs follow-up) and to examine whether the Intervention and BAU groups differed in terms of their mean change over time by examining the Time (baseline Vs follow-up) by Condition (BAU Vs Intervention) interaction. In all cases we included the children's language status (monolingual English vs DLL) as a between-group factor and age of baseline assessment as a covariate. We report effect sizes[2] and highlight these in conjunction with p levels. To examine the effects of the intervention on narrative skills we ran four multivariate analysis of covariance (MANCOVAs) with language status, area, gender and condition as predictors, Bus Story variables as the outcome and age at follow-up as the covariate. Language status was included as an additional covariate when analysing by area given the significantly higher proportion of children who were DLL in the London group.

Analyses are based on raw data. For the BAS-3 subtests, children were administered a specific set of items based on their age. Due to the variability in item sets, the raw scores are converted into ability scores to allow for comparison between scales. Ability scores not only demonstrate the number of items answered but also the difficulty of those items. Ability scores are calculated using the Rasch Model (a form of factor analysis) which provides statistically reliable scales that can be compared across different sets of items, and also allows for measurement of change over time. Means (*SD*s) for baseline and follow-up assessments by intervention group and language status are presented in Appendix G1-G9.

3.11.1.1 Bus Story

The Bus Story audio recordings were transcribed by a research assistant unaware of the study design and purpose. CF and another research assistant not involved in the data collection coded the transcripts following the guidelines provided (Renfrew, 1997). Two schools (n =

[2] Effect sizes in this report adhere to the following definitions (Cohen, 1988): Small: d = .02; Medium: d = .05; Large: d = .08.

21 transcripts) were selected for an initial calibration exercise. Scoring discrepancies were discussed, decisions were documented and a further three schools (*n* = 38 transcripts) were coded using these refined guidelines. Again, discrepancies were discussed until agreement was reached. A further four schools were randomly selected (*n* = 33 transcripts) to determine inter-rater reliability. A mean-rating (k = 2), absolute-agreement, two-way mixed effects model was used to calculate intra-class correlations (ICC) in SPSS. Inter-rater reliability was 'excellent' for both Information (ICC = .98 [95% CI = .96-.99]) and Sentence Length (ICC = .95 [95% CI = .91-.97]) (Cicchetti, 1994). Due to the young age group of the sample only a small minority of the total sample (13%) produced composite sentences, therefore no Sentence Complexity score was computed for reliability. The remaining transcripts were coded independently with coders blind to the condition of each school.

3.11.2 Practice Measures

To evaluate the impact of the *Talking Time*© intervention on practitioner use of language supporting strategies, we focused on 'Storm Whale Reading' recordings. Of the three recording activities, this text-and-picture book reading activity was considered the most relevant, due to the importance of shared reading in developing children's oral language, and the most representative of typical classroom practice. All recordings with matched baseline and follow-up submissions were transcribed in full by a trained research assistant. Practitioner and child speech was segmented into C-units following guidance from the Systematic Analysis of Language Transcripts (SALT) Standard Transcription Conventions (Systematic Analysis of Language Transcripts, 2020). Each C-unit was defined as an independent clause with its modifiers. For consistency, the accuracy of C-unit segmentation of transcripts was checked by a single member of the research team prior to coding. Following a review of existing research literature, a bespoke coding frame was developed (Table 3). More details of the rationale and provenance underpinning the selection of these variables can be found in Appendix E1.

Table 3: Audio recording variables and codes

Variable	Codes	Definition
Pragmatic function	a) Extra textual utterances b) Reading utterances c) Child utterances d) Unrelated utterances	Narrative utterances relating to the text. Direct reading of the text. Child speech. Adult utterances related to behaviour management and praise unrelated to responses to the text.
Conversations	a) Conversations	A back and fore exchange on a topic with a minimum of two turns beginning with an

	b) Conversational turns	initiating prompt and ending where there was no semantic link between utterances, or where the practitioner: • returned to reading directly from the text; • broke off to manage behavior; or • repeated a child utterance to end the turn. Changes in speaker within each conversation following the initiating prompt.
Adult prompts	a) Open prompts	Adult question or comment seeking a child response for which there is no single predetermined answer and for which an adequate response would be more than one word.
	b) Closed prompts	Adult question or comment seeking a child response which allow for one or a limited number of acceptable or pre-determined responses.

Transcripts were assigned anonymised codes to ensure that coders were blind to condition. Four transcripts were selected for training and familiarisation. JC, CF and KD coded the transcripts independently using the first version of the coding framework. Differences were discussed and refinements were made to the coding framework to support consistency of decision making. CF and KD coded eight further transcripts and reliability was calculated for all key codes. A mean-rating (k = 2), absolute-agreement, two-way mixed effects model was used to calculate intra-class correlations (ICC) in SPSS. Inter-rater reliability was 'excellent' for all codes, ranging from .923 to .996 (Cicchetti, 1994). The remaining matched transcripts were then divided by area (London vs Northeast) between the two coders.

The submitted recordings were coded for length (minutes and seconds; MM:SS) and ranged from 04:08 to 23:44. To control for variation in recording length, proportion scores were created for count variables to allow accurate comparisons between practitioners. Prompt variables were analysed as a proportion of adult extra textual utterances and pragmatic function categories as a proportion of total utterances. Descriptive statistics for coded and derived variables are reported in Appendix E2.

In line with the analysis of child data a two-by-two repeated measures analysis of variance (ANOVA) with Time (baseline Vs follow-up) as the repeated measure and Condition (BAU vs Intervention) as the between groups measure was used for each variable. Variables were inspected for normality and, for the proportion scores, the distribution of residuals was also checked (Chen et al., 2017). Where indications of skewed data were

detected, transformations (Log-10 and square root) were applied to normalise distributions. Where transformations were undertaken, non-parametric Quade's ANCOVAs considering differences between groups at follow up after controlling for baseline measures were run alongside the parametric tests as a confirmatory measure.

3.11.3 Parent/Carer Reported Data

LUI-6 English and Home data were scored by one member of the research team. Incomplete LUI forms were scored as missing data following LUI scoring guidelines. SDQ data were automatically scored using formulae calculating subscale and total scores. Data from both questionnaires were analysed using a two-by-two repeated measures ANOVA with Time (baseline Vs follow-up) as the repeated measure and Condition (BAU vs Intervention) as the between groups measure. All variables were tested for normality and any skewed variables were transformed using square root transformation.

3.11.4 Implementation and Process Evaluation

Interviews were recorded and automatically transcribed. Transcripts were checked against recordings and corrected where necessary. A thematic coding frame (Implementation Annex Appendix 4) was developed based on the dimensions and factors of implementation detailed in the Education Endowment Foundation's IPE guidance (Humphrey et al., 2016) and wider literature (e.g. (e.g., Neugebauer et al., 2021). Initially, one transcript was coded independently by two researchers and discrepancies discussed to support framework refinement. Subsequent transcripts were coded using Nvivo 12 by one researcher. Transcripts from two schools (one in each area) were independently coded by an external coder, resulting in a mean Cohen's kappa statistic of 0.75[3], indicating substantial agreement (Landis & Koch, 1977). A small number of additional codes emerged during coding and were added to the framework. Qualitative survey responses were coded using the same framework. Means of all rated survey items were calculated to support descriptive analysis.

Weekly session logs were collated and analysed at school and child level. At school level, the amount of delivery and proportion of intended delivery were determined. To enhance accuracy, data were adjusted to account for missing records. Where it could be established (e.g., via interview) that sessions were delivered but not recorded, the number of intended sessions was reduced accordingly, and adjusted delivery figures calculated. If established that a record was missing because no sessions were delivered that week, this was

[3] Before running the coding comparison, instances where one coder had included an interviewer question/contribution within their coded text segment, but the other had not, were identified and aligned.

recorded as 'no provision'. Each child's exposure to the programme was also determined, adjusted for missing records as needed. Since the four schools which dropped out prior to the end ('leavers') did not provide records, a broader categorisation was created to assess fidelity across all 18 schools that began the programme. Schools were categorised into four groups (≥ 100%; ≥ 80%; ≥ 50%;< 50%) dependent on the proportion of intended delivery and child exposure achieved (adjusted for missing weeks). Delivery of <50% was assumed for leaver schools, based on mentor feedback.

4. Impact evaluation

The outcomes of the trial are discussed in the following five sections:

1. Sample and attrition
2. Baseline data - *demonstrating the language abilities, nursery practice and parent perceptions of child language and behaviour before the intervention started*;
3. Effect of language status - *examining the difference between monolingual English speakers and children who are DLL;*
4. Growth over time - *demonstrating the difference between data collected at baseline and follow-up (approximately seven months later;*
5. Intervention effects - *showing the difference between the intervention group and the BAU group, with significant gains for the intervention group demonstrating the success of the intervention.*

Within each section we present child measures, practice measures and parent/carer reported measures.

4.1 Sample and attrition

4.1.1 Child Measures

Parental/carer consent was obtained for 584 children (London *n* = 375; Northeast *n* = 209) from 38 schools. Of those children, 526 completed assessments at baseline (London *n* = 327; Northeast *n* = 199). Fifty-eight children (10%) (London *n* = 48; Northeast *n* = 10) did not complete assessments at baseline. Twenty-three of those children were not eligible because either the parent or teacher reported significant special educational needs (SEN) (*n* = 13), the child was too young (*n* = 4), the consent was incomplete or withdrawn (*n* = 3), or the child no longer attended the nursery class (*n* = 3). Children who did not complete the baseline assessment were more likely to be younger (U = 17862.00, *p* = .007, 95% CI = -3.00, .00), and more likely to attend nursery part-time (X^2 (1, *n* = 449) = 4.29, *p* = .038, Cramer's V = .098). At follow-up, 45 of these children did not complete assessments (London *n* = 21; Northeast *n* = 24) (see Figure 2 for details). A further thirty-five children were lost to follow-

up because their school dropped out of the study before the intervention began (*n* = 24 children) or after randomisation (*n* = 11 children). This resulted in a final matched sample of 446 children (London *n* = 278; Northeast *n* = 168).

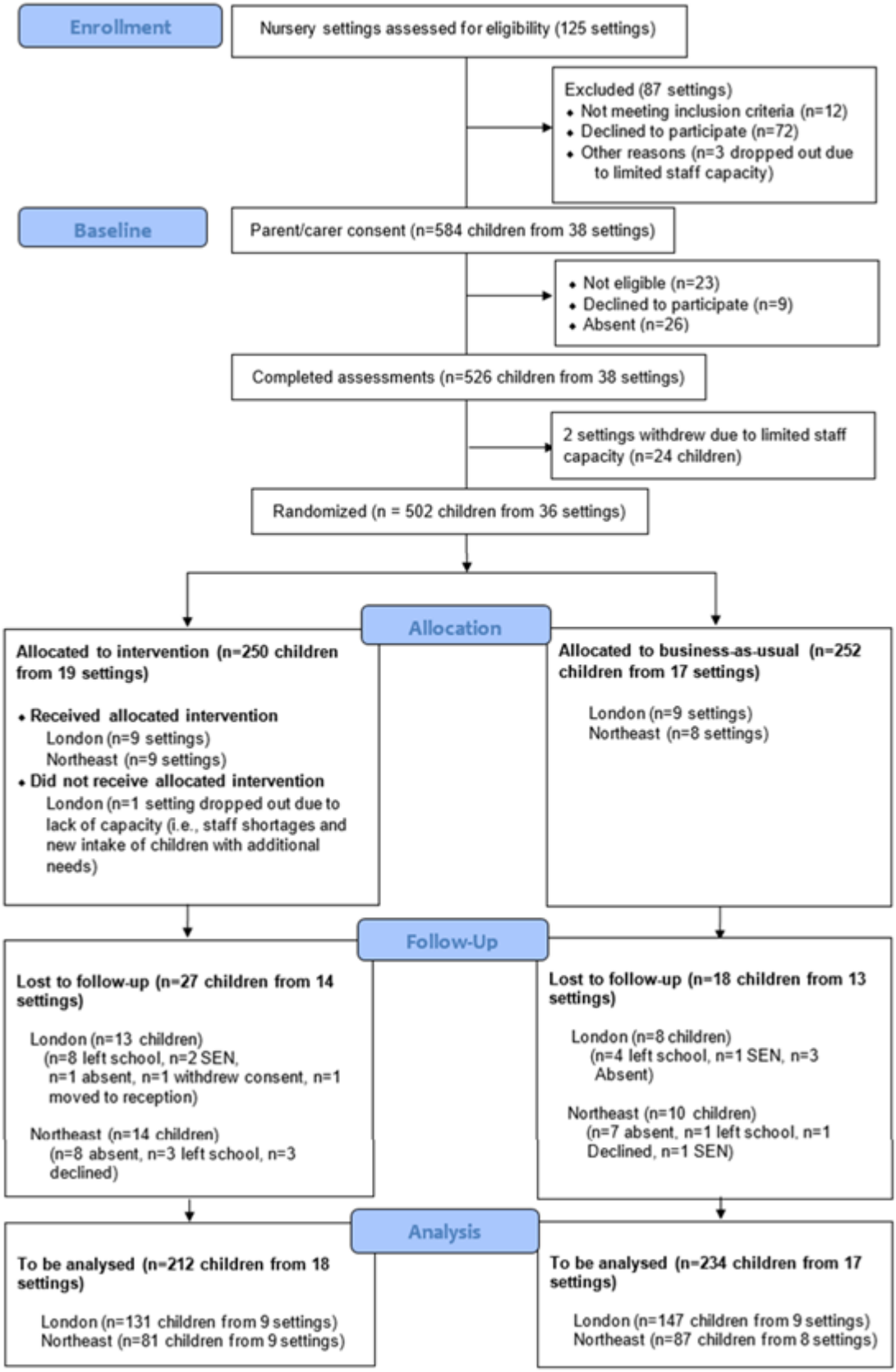

Figure 2: CONSORT flow diagram.

4.1.2 Practice Measures

Of the 103 practitioners invited to register on the secure server (REDCap), 61 did so, 51 of whom were *Talking Time*© leads (see Appendix F2 for demographics). Forty-three practitioners submitted baseline Storm Whale Reading audio recordings, twenty-nine of whom also submitted follow up recordings (see Appendix F3 for practitioner flow throughout the study). Chi-squared tests revealed practitioners in London were more likely to submit recordings at both time points than those in the Northeast (χ^2 (1,61) = 5.335, p = .021) There were no other significant differences between practitioners who registered on the secure server and those who submitted baseline and follow up audio recordings in relation to role, condition, highest level of qualification or years of experience (all ps > .05). Following the exclusion of three audio recordings due to recording errors, twenty-six matched baseline and follow up recordings were available for coding (n = 10 BAU, n = 16 intervention).

4.1.3 Parent/carer Reported Data

Of the 584 parents/carers recruited to the study, 172 (29%) returned the LUI-English at baseline (Intervention n = 101, BAU n = 71). Eighty-five parents had matched baseline and follow-up LUI-English data. Fifteen of these matched LUIs were incomplete with missing items, resulting in a final sample of 70 parents/carers (12%) with complete LUI-English baseline and follow-up data (Intervention n = 43, BAU n = 27). Chi-square analysis revealed there were no significant differences between those who did and did not return LUI-English, based on parent education, language status, parent mental health, child language concern or child gender (all ps > .05). Seventy-two parents/carers [23% of the total dual language learner (DLL) parents/carers] returned the LUI-Home at baseline (Intervention n = 39, BAU n = 33). Of those, 28 parents/carers (9%) had matched baseline and follow-up LUI-Home data (Intervention n = 18. BAU n = 10).

In total, 161 parents/carers (28% of recruited sample) returned the SDQ at baseline (Intervention n = 100, BAU n = 61). Eighty-three parents/carers (14% total sample) had matched baseline and follow-up SDQ data (Intervention n = 56, BAU n = 27). Chi-square analysis indicated parents were more likely to return the SDQ if they did not have a concern about their child's language ($\chi^2(1) = 4.16$, p = .042).

4.2 Baseline data

4.2.1 Child Measures

At baseline child language was below average compared to norms. Monolingual English-speaking children performed better than Dual Language Learners (DLLs) on all tasks except the Non-word Repetition task:

Baseline and follow-up assessments were completed in 35 schools with 446 children (London n = 278; Northeast n = 168). Demographic details of the sample are presented in Appendix F4. The sample ranged in age from 3 years 0 months to 4 years 8 months (M_{age} = 3;8) at baseline. Half of the sample were girls, over a third of the sample were first born and 57% attended nursery part-time. Parents reported low levels of reading to their children with less than half of the sample read to every day. Approximately one in five reported concerns about their child's language development.

The two locations varied in level of deprivation, language spoken at home and ethnicity, with schools in the Northeast being significantly more deprived than schools in London. Over three-quarters of the London sample were DLLs compared to less than one-fifth of the sample in the Northeast and a significant proportion of London parents/carers come from a Bangladeshi background while the majority of parents/carers in the Northeast were White British. Similarly, more of the London sample had lived outside of the UK since their child was born compared to the Northeast sample. Most of the sample had completed post-secondary education but this varied between locations, with more respondents in London achieving university education. Fifty-eight percent of the respondents in the sample were employed at baseline. The Office for National Statistics (ONS) figures from the same time-point show the UK population employment rate at 76% (Watson, 2021). The self-reported literacy abilities of the parents/carers were similar between locations, with around 10% stating that they had difficulty reading children's books and completing forms. Finally, over a third of respondents in the Northeast (35%) reported mental health difficulties, a significantly higher proportion than those in London (10%). Overall, 19% of the sample reported speaking to a doctor about their mental health, in line with the17% of adults who report experiencing a common mental disorder at the national level (McManus et al., 2016).

4.2.1.1 Differences between groups at baseline

Appendices G1 and G2 show the raw child outcome data by condition and language status at baseline and follow-up. Appendix G3 shows statistical comparisons of language status. At baseline, the monolingual English children performed significantly better on most of the language tasks except the Non-word Repetition task where children who were Dual Language Learners (DLL) scored significantly higher. Additionally, the children who were DLL knew

significantly more target object words at baseline compared to the children who spoke only English at home. There was no significant difference on Sentence Repetition or on the non-verbal tasks. There were no significant differences in language ability based on gender or deprivation level.

There were no demographic differences between the Intervention and the BAU groups (see Appendix G4 for details). There was a significant difference between groups at baseline for the Target Object Naming task ($F(1, 353) = 4.22$, $p = .041$, $\eta_p^2 = .012$), with children in the BAU group scoring significantly higher than children in the Intervention group before randomisation. There were no other significant group differences before randomisation.

At baseline, the *Talking Time* sample as a whole performed significantly below average compared to expected population norms on most of the tasks with medium to large effect sizes, except for the GAPS Nonword Repetition ($t(372) = .70$, $p = .486$, Cohen's $d = .04$) where children were within the average range (see Appendix G5 for details). The sample also performed significantly below population norms on all three of the domains of the Bus Story (Information: $t(405) = -35.12$, $p < .001$, Cohen's $d = -1.74$; Sentence Length: $t(279) = -12.22$, $p < .001$, Cohen's $d = -.73$; Sentence Complexity: $t(279) = -34.40$, $p < .001$, Cohen's $d = -2.06$) tested at follow-up.

4.2.2 Practice Measures

At baseline nursery staff demonstrated a range of practices More closed prompts than open prompts were used at baseline:

The 26 matched recordings were submitted by teachers, teaching assistants, nursery nurses and early years leads whose years of experience ranged from 3 to 40 years. The full range of levels of qualification (Level 2 - Level7/8) were represented (see Appendix E4 for more details). There were no significant differences between Intervention and BAU groups in terms of qualification, role, working pattern or years of experience (all *p*s > .05). Group level information in respect of the child participants is set out in Appendix E5.

At the whole group level, the length (mm: ss) of baseline recordings varied between practitioners but on average were around 10 minutes long (M = 10:08, SD = 03:52) consisting of over 200 utterances (M = 260, SD = 144). On average these utterances were made up of 30% child utterances, 15% adult reading utterances (direct reading of the text) and 53% extra textual utterances (conversational comments or questions based on the story), with the remainder of utterances unrelated to the topic. Of the adult extra-textual utterances, an

average of 32% (*SD* = .11) were closed prompts and 12% (*SD* = .05) were open prompts. The mean number of conversations per recording was 18 (*SD* = 8.7) and on average these conversations contained 5.36 (*SD* = 1.68) conversational turns. Details of findings by condition are provided in Appendix E3. There were no significant differences between groups at baseline in respect of any of the coded variables, except for the proportion of closed prompts which was significantly higher in the BAU group (t(24) = 2.496, *p* = .020).

4.2.3 Parent/carer Reported Data

Parent/carers reported strong child language skills overall, but children's use of home language was slightly lower than use of English. A significant proportion of children received 'High' or 'Very High' ratings of challenging behaviours compared to the UK population as a whole:

4.2.3.1 Language: LUI-6

Parent Report of Children's English Language Use: LUI-6-English

Intervention and BAU groups were comparable in Total LUI-English scores at baseline (*p* = .942). Mean LUI scores indicated that for most subscales children scored above 80% for proportion of items identified at baseline (range = 3-100%), with the exception of subscale N, 'Building longer sentences and stories', where children scored below 70%. Appendix G11 displays the proportion of LUI-English items identified as correct for both groups at baseline and follow-up.

Parent Report of Children's Home Language Use: LUI-6 Home

Total LUI-Home scores for Intervention and BAU groups were comparable at baseline (*p* = .226). Mean scores indicate variability in proportion of items correct between LUI-Home subscales (range = 4-100%), and large standard deviations indicate variability between children, i.e., some were using their home language well while others were less competent using their home language. Appendix G12 displays the proportion of correctly identified LUI-Home items for both groups at baseline and follow-up.

4.2.3.2 Behaviour – SDQ

Mean SDQ scores were similar between BAU and Intervention groups at baseline, and indicate children were reported to have the greatest difficulties with hyperactive behaviour. The proportion of children in BAU and Intervention groups scoring in each SDQ category (indicating degree of difficulties) were similar and Chi-square analysis indicated there were no significant group differences at baseline ($\chi^2(3) = 3.67$, $p > .05$). Appendix G13 and G14 display the descriptive statistics for the SDQ and categorical proportions for both groups at baseline and follow-up.

4.2.3.2.1 Population differences

Scores indicated that the majority of children in both groups scored in the 'close to average' category for total difficulties score at both time points (score < 12). Figure 3 displays the proportion of the Talking Time sample falling into each SDQ category for degree of difficulty compared to the wider UK population norms (scores of 0-12 'Close to average'; 13-15 'Slightly raised; 16-18 'High' and 19-40 'Very high'). More children in the Talking Time sample scored in the 'High' and 'Very high' categories at both time points compared to the UK population. Differences in proportions between the Talking Time sample and UK population were statistically significant at both time points (Baseline; $\chi^2(3) = 23.84$, $p < .001$; Follow-up; $\chi^2(3) = 12.11$, $p = .007$).

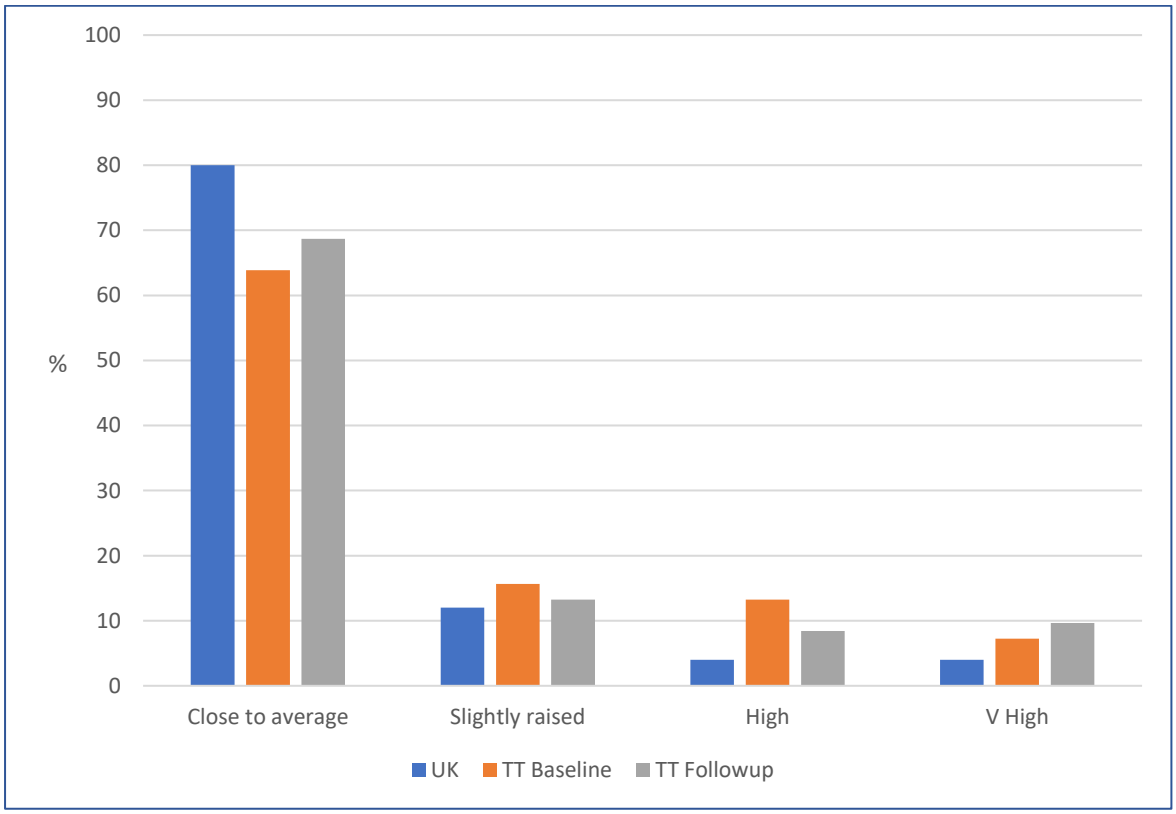

Figure 3: Strengths and Difficulties Questionnaire (SDQ) Total Score cut-offs for Talking Time sample and UK norms.

4.3 Effect of language status

In this section, the main effect of language status (monolingual English vs Dual Language Learners (DLL)) is analysed to examine whether children differ in their performance due to the language spoken at home. See Appendix G for more details.

4.3.1 Child Measures

Monolingual children performed better on all language tasks than the children who were DLL, except for the Non-word Repetition task. Although monolingual children achieved higher scores at post-test on the majority of oral language measures, children who were DLL made more progress over time on verbal comprehension and expressive vocabulary:

There was a statistically significant main effect for language status across the tasks, with monolingual English-speaking children scoring significantly higher than the children who were DLL on the Sentence Repetition task ($F(1,376) = 4.62$, $p = .032$, $\eta_p^2 = .012$); Naming Vocabulary task ($F(1,394) = 69.99$, $p < .001$, $\eta_p^2 = .151$); the Verbal Comprehension task ($F(1,397) = 56.04$, $p < .001$, $\eta_p^2 = .124$); the QUILS Vocabulary area ($F(1,325 = 29.34$, $p < .001$, $\eta_p^2 = .083$); Syntax area ($F(1,325 = 19.99$, $p < .001$, $\eta_p^2 = .058$) and Process area ($F(1,325 = 13.08$, $p < .001$, $\eta_p^2 = .039$). The opposite effect was found for Non-word Repetition with children who were DLL scoring significantly higher than monolingual English children ($F(1,380 = 13.15$, $p < .001$, $\eta_p^2 = .033$).

For the bespoke measures there was a significant effect of language status with monolingual English children scoring higher on control action words ($F(1,389 = 25.58$, $p < .001$, $\eta_p^2 = .062$), target object words ($F(1,392 = 12.69$, $p < .001$, $\eta_p^2 = .031$) and target action words ($F(1,389) = 39.73$ $p < .001$, $\eta_p^2 = .093$). There was no effect of language status on control object words ($p > .05$).

4.3.1.1 Change over time by language status

There was a significant interaction between language status and time on the Verbal Comprehension measure ($F(1,397) = 8.97$, $p = .003$, $\eta_p^2 = .022$), reflecting the greater increase in performance over time by the children who were DLL (see Figure 4).

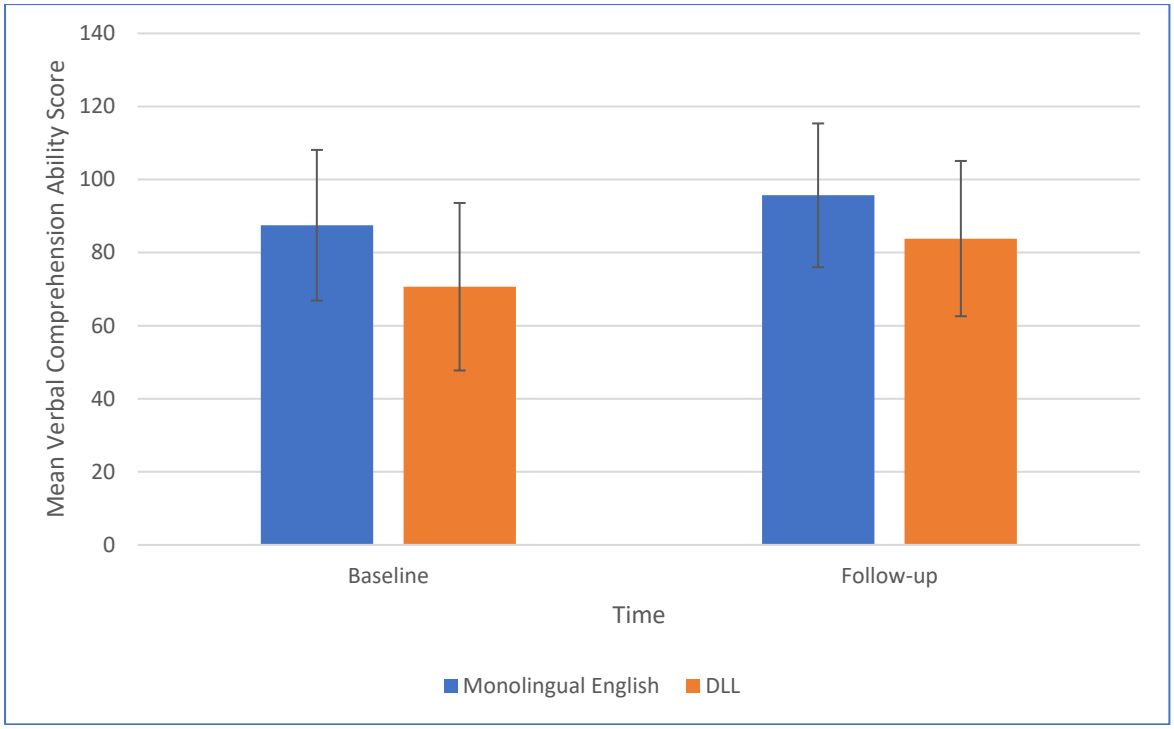

Figure 4: **Interaction between time and language status for Verbal Comprehension.**
Note: Error bars denote SDs.

There was a significant interaction between language status and time on the Naming Vocabulary measure (F(1,394) = 15,40, p <.001, η_p^2 = .038), reflecting a greater increase in performance over time by the children who were DLL (see Figure 5).

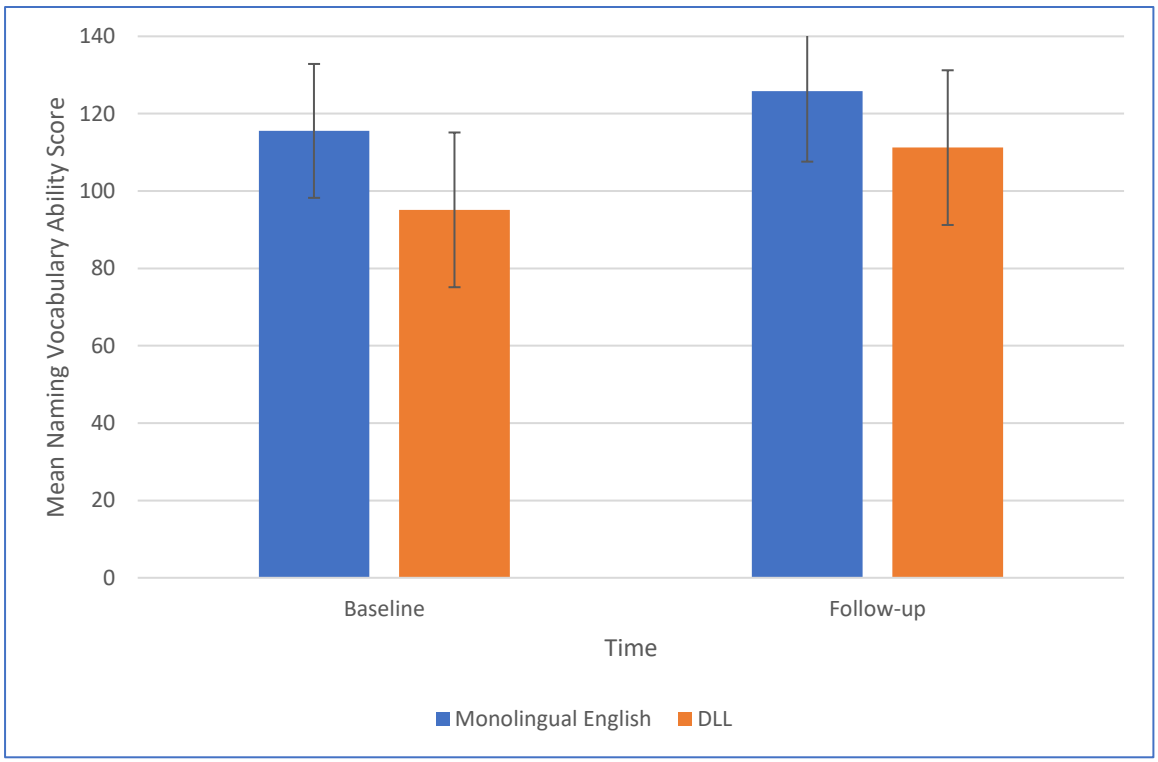

Figure 5: **Interaction between time and language status for Naming Vocabulary.**
Note: Error bars denote SDs.

4.3.2 Parent/carer Reported Data

There was no main effect of language status on Total Difficulties score ($p > .05$) and no interaction effect of language status over time for Total Difficulties score, or any of the individual subscales, as all comparisons were non-significant ($p > .05$).

4.4 Growth over time

In this section, the main effect of time is examined (i.e., whether there were significant changes in performance in the approximately seven months between baseline and follow-up data collection). See Appendix G for details. Findings for child and practice measures are also summarised in Infographics 1 (p. 49) and 2 (p. 52), respectively.

4.4.1. Child Measures

There was significant growth over time for some of the standardised language tasks but not for the control or bespoke measures:

Control Measures

Children had significantly higher scores at follow-up on the Non-word Repetition task ($F(1,380 = 7.85$, $p < .001$, $\eta_p^2 = .020$) and the Matrices task ($F(1,388 = 10.75$, $p < .001$, $\eta_p^2 = .027$). There was no significant improvement over time on Picture Similarities ($p > .05$).

Standardised Language Measures

Children had significantly higher scores at follow-up on the Sentence Repetition task ($F(1,376) = 4.37$, $p = .037$, $\eta_p^2 = .011$); the Naming Vocabulary task ($F(1,394) = 4.99$, $p = .026$, $\eta_p^2 = .013$); the Verbal Comprehension task ($F(1,397) = 11.76$, $p < .001$, $\eta_p^2 = .029$) and the QUILS Vocabulary area ($F(1,325 = 5.75$, $p = .017$, $\eta_p^2 =. 017$). There was no significant improvement over time on the QUILS areas of Syntax or Process (ps > .05).

Bespoke Vocabulary Items

There was no significant improvement over time for the Object or Action target words. Similarly, there was no significant improvement over time for Object or Action control words (all ps > .05).

4.4.2 Practice Measures

At follow-up, audio-recordings were longer and contained a higher proportion of child utterances (i.e. children contributed more) than at baseline. Adults also read less directly from the text at follow-up:

Length of recordings

There was a significant main effect of time on the duration of recordings ($F(1.24) = 11.72$, $p = .002$, $\eta p^2 = .328$) and the number of utterances ($F(1,24) = 14.912$, $p < .001$, $\eta p^2 = .383$), with an increase in both groups (intervention and BAU) at follow-up.

Utterance types

There was a significant main effect of time on the proportion of child utterances ($F(1,24 = 36.08$, $p < .001$, $\eta p^2 = .601$) which increased at follow-up. There was a significant main effect of time on the proportion of adult reading-from-the-text utterances ($F(1,24) = 36.60$, $p < .001$, $\eta p^2 = .604$), which decreased from baseline to follow-up. There was no effect of time on the proportion of adult extra-textual utterances ($p > .05$).

Conversations

There was a significant main effect of time on average conversational turns ($F(1,24) = 25.01$, $p < .001$, $\eta p^2 = .510$ with an increase at follow-up. The total number of conversations over time approached significance ($p = .050$).

Prompts

There was no effect of time on the proportion of extra textual utterances which were closed or open prompts (both comparisons were non-significant, $p > .05$).

4.4.3 Parent/carer reported measures

Parent/carer ratings of child language and behaviour did not change over time:

Parent/carer Reported Language

There was no significant effect of time on LUI-English total score or LUI-Home total score (all *p*s > .05), suggesting there was no change in parent-reported child language between baseline and follow-up.

Parent/carer Reported Behaviour

There was no effect of time on the SDQ Total Difficulties score ($p = .864$), indicating no significant change in challenging behaviours from baseline to follow-up. This was also true for the individual subscales of Emotional difficulties, Conduct problems, Hyperactivity, Peer difficulties and Prosocial behaviour where all comparisons were non-significant ($p > .05$).

4.5 Intervention effects

In this section, significant interactions between time (Baseline vs Follow-up) and condition (Intervention vs BAU) demonstrate the impact of the intervention (i.e., assessing whether the intervention group made greater progress over time than the BAU group). At the child level we predicted significant intervention effects for the targeted variables of vocabulary, comprehension and narrative skills but not the control variables of non-verbal abilities and non-target vocabulary. At the practice level we expected significant changes in adult-child interactions for those in the intervention group but not the BAU group. We did not expect any effect of the intervention on parent/carer reported data. See Appendix G for details of child

and measures parent/carer report. Intervention effects are summarised in Infographics 1 (p. 49) and 2 (p. 52), respectively.

4.5.1 Child Measures

The intervention had a significant effect on the bespoke vocabulary measure (both target object and action words), with children in the intervention group scoring significantly higher. However, there was no significant impact of the intervention on standardised language measures, although a trend was evident for sentence repetition. As expected, no intervention effects were identified for the control measures:

Control Measures

There was no intervention effect for Matrices or Picture Similarities (*p*s > .05). There was no intervention effect on Non-word Repetition (p = .404). However, as shown in Figure 6, there was a significant three-way interaction between time, condition and language status (F(1, 380) = 5.20, p = .023, η_p^2 = .014): children who were DLL in the intervention condition showed less growth over time than the children who were DLL in the BAU condition.

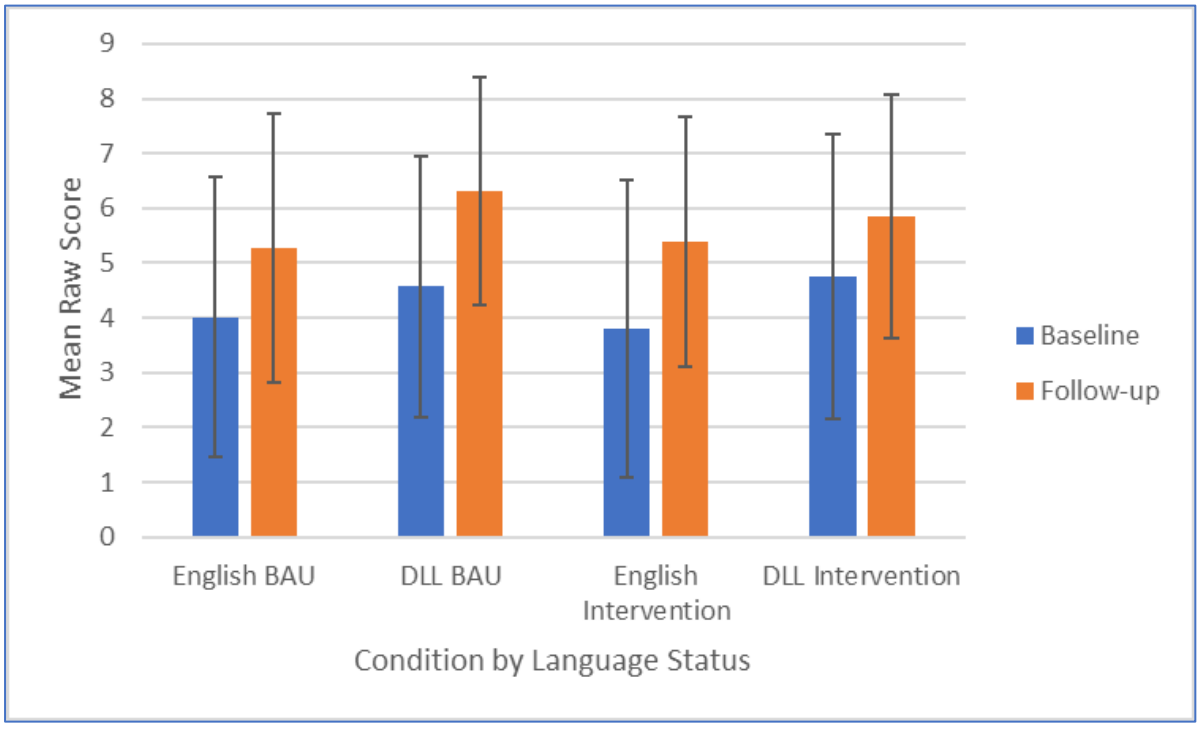

Figure 6: **Interaction between time, condition and language status for Non-word Repetition.**

Standardised Language Measures

The intervention effect for Sentence Repetition performance fell just below statistical significance (F(1,376) = 3.17, p = .076, η_p^2 = .008) suggesting that an effect may have been identified in a larger sample. The intervention had no significant impact on Naming

Vocabulary scores, Verbal Comprehension scores, the QUILS assessments, or Bus Story scores (all *ps* > .05).

Bespoke Vocabulary Measures

There was a significant effect of intervention on the target vocabulary embedded within the intervention programme, with children in the intervention group improving significantly more than children in the BAU group on both Object words ($F(1,392) = 6.38$, $p = .012$, $\eta_p^2 = .016$) and Action words ($F(1,379) = 9.86$ $p < .001$, $\eta_p^2 = .025$). There was also a three-way interaction between time, condition and language status $F(1,389) = 6.33$ $p = .012$, $\eta_p^2 = .016$) which reflected a greater growth in target Action words for the monolingual intervention cohort as shown in Figure 7. As predicted, the intervention had no effect on the control items for Object or Action words (*p*s > .05).

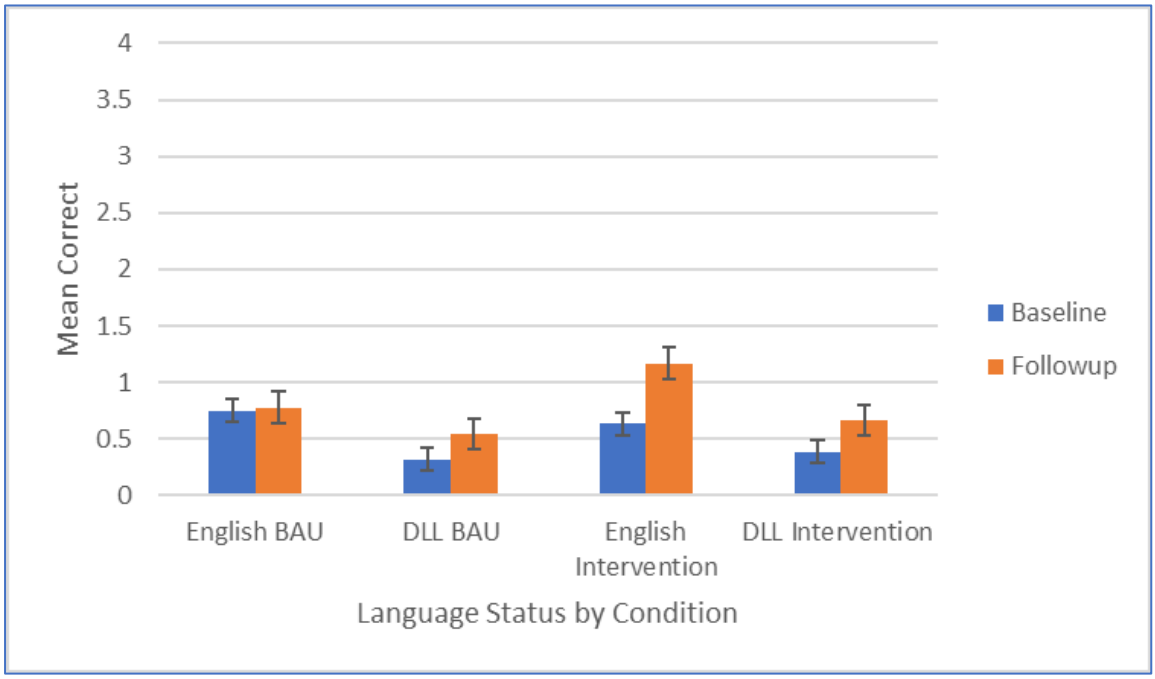

Figure 7: **Interaction between time, condition and language status for target Action words.**

Infographic 1 summarises the main effects and interactions of time, condition and language status on the child outcomes of verbal and non-verbal abilities.

	Outcome Measure	Impact of DLL	Growth over time (approx. 7 months)	Intervention effect
Control Measures	BAS-3 Picture Similarities	✗	✗	✗
	BAS-3 Matrices	✗	✓	✗
	GAPS Non-word Repetition	✓ Higher scores in DLL group	✓	✗
Structural Language Ability	GAPS Sentence Repetition	✓	✓	Non-significant trend in favour of intervention
	BAS-3 Verbal Comprehension	✓	✓	✗
	BAS-3 Naming Vocabulary	✓	✓	✗
	QUILS Vocabulary	✓	✓	✗
	QUILS Syntax	✓	✗	✗
	QUILS Processing	✓	✗	✗
Bespoke Vocabulary Items	Control Object words	✗	✗	✗
	Control Action words	✓	✗	✗
	Target Object words	✓	✗	✓
	Target Action words	✓	✗	✓

Infographic 1: Summary of intervention impact on child measures
Note: The impact of DLL reflects a higher score for monolingual English speakers unless otherwise stated.

4.5.2 Practice Measures

The intervention significantly impacted the quality of adult-child book sharing interactions. Practitioners in the intervention group reduced their direct reading from the text over time and used a higher proportion of language-eliciting prompts, compared with the BAU group. Children in the intervention group contributed more and the average number of conversational turns increased, suggesting longer back and forth exchanges between the adults and children:

Length of recordings

There was a significant intervention effect on the length of recordings ($F(1,24) = 9.12$, $p = .006$, $\eta p^2 = .275$), with the intervention group submitting longer recordings than the BAU group at follow-up. In line with the longer duration of recorded activities, the intervention also positively impacted utterances ($F(1,24) = 4.80$, $p = .038$, $\eta p^2 = .167$), with a greater number of utterances in intervention recordings at follow-up.

Utterance types

There was a significant intervention effect on child utterances ($F(1,24 = 4.68$, $p = .041$, $\eta p^2 = .163$), with a larger growth in the proportion of child utterances in the intervention group. There was a significant main effect of condition on the proportion of adult direct text-reading utterances ($F(1,24) = 11.05$, $p = .003$, $\eta p^2 = .315$) and a significant intervention effect ($F(1,24) = 17.06$, $p < .001$, $\eta p^2 = .415$), with less direct reading from the text in the intervention group at follow-up. Due to difficulties correcting the skewness of these data, a non-parametric ANCOVA was also employed to ensure robustness. The finding was confirmed with a significant difference between groups after controlling for the proportion of adult reading utterances at follow-up ($F(1,24) = 18.66$, $p < .001$, $\eta p^2 = .437$).

There was no significant intervention effect on the proportion of adult extra-textual utterances ($p = .725$).

Conversations

There was a significant intervention effect on average conversational turns ($F(1,24) = 9.60$, $p = .005$, $\eta p^2 = .286$), with a greater increase in the intervention group at follow-up than the BAU group. There was no effect of the intervention on the total number of conversations ($p = .096$).

Prompts

The intervention significantly impacted the proportion of adult extra-textual utterances which were prompts ($F(1,24) = 9.04$, $p = .006$, $\eta p^2 = .273$), with an increase in the intervention group and a reduction in the control group.

When analysed by type, there was a significant interaction effect on closed prompts ($F(1,24) = 8.10$, $p = .009$, $\eta p^2 = .252$), with an increase in the intervention group and a reduction in the control group. The proportion of open prompts was not significantly impacted by the intervention ($p = .260$) (see Table E3 in Appendix).

	Variable	Growth over time (approx. 7 months)	Intervention effect
Length	Duration (minutes:seconds)	✓	✓
	Number of utterances	✓	✓
Utterance type	Proportion adult extra textual	✗	✗
	Proportion adult reading direct from text (decrease)	✓	✓
	Proportion child utterances	✓	✓
Conversations	Total number of conversations	✗	✗
	Average turns in each conversation	✓	✓
Prompts	Proportion extra textual utterances which were prompts	✗	✓
	...which were closed prompts	✗	✓
	...which were open prompts	✗	✗

Infographic 2: Summary of intervention impact on practice measures.

4.5.3 Parent/carer Reported Data

Parents/carers in the intervention group reported significantly greater improvement in their child's home (non-English) language than parents in the BAU group. No effect was identified for children's use of English as reported by parents/carers. There was no intervention effect on parent/carer report of behaviour:

4.5.3.1 Parent/carer Reported Language

The LUI provided two measures, one for use of English and one for use of Home Language. There was no significant main effect of the intervention on the LUI-English total score (p = .785). Mean proportion of identified items for each LUI scale at follow-up are displayed in Figure 8.

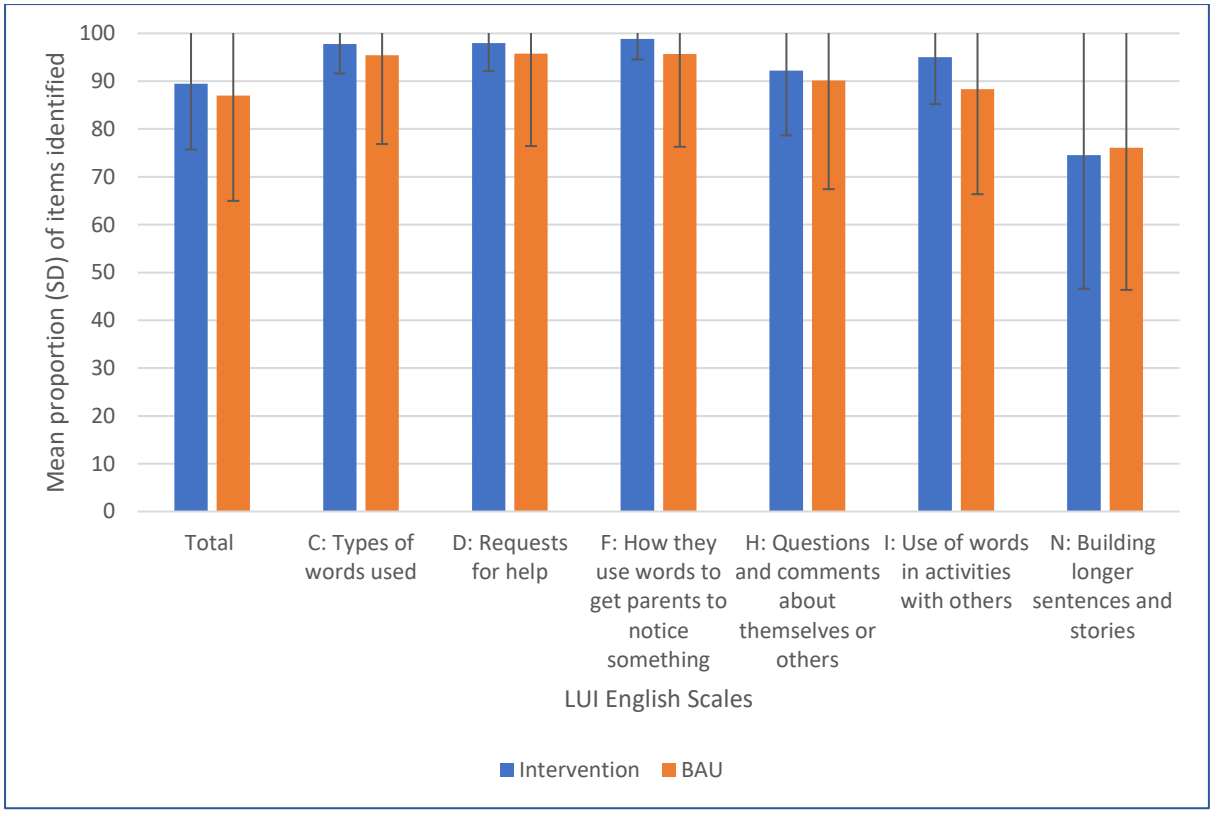

Figure 8: Parent/carer report of English language ability at follow-up.

There was a significant intervention effect for LUI-Home total score (F(1) = 4.59, p = .044 ηp^2 = .179), with a greater increase over time in the number of identified items in the Intervention group compared to the BAU group (Figure 9).

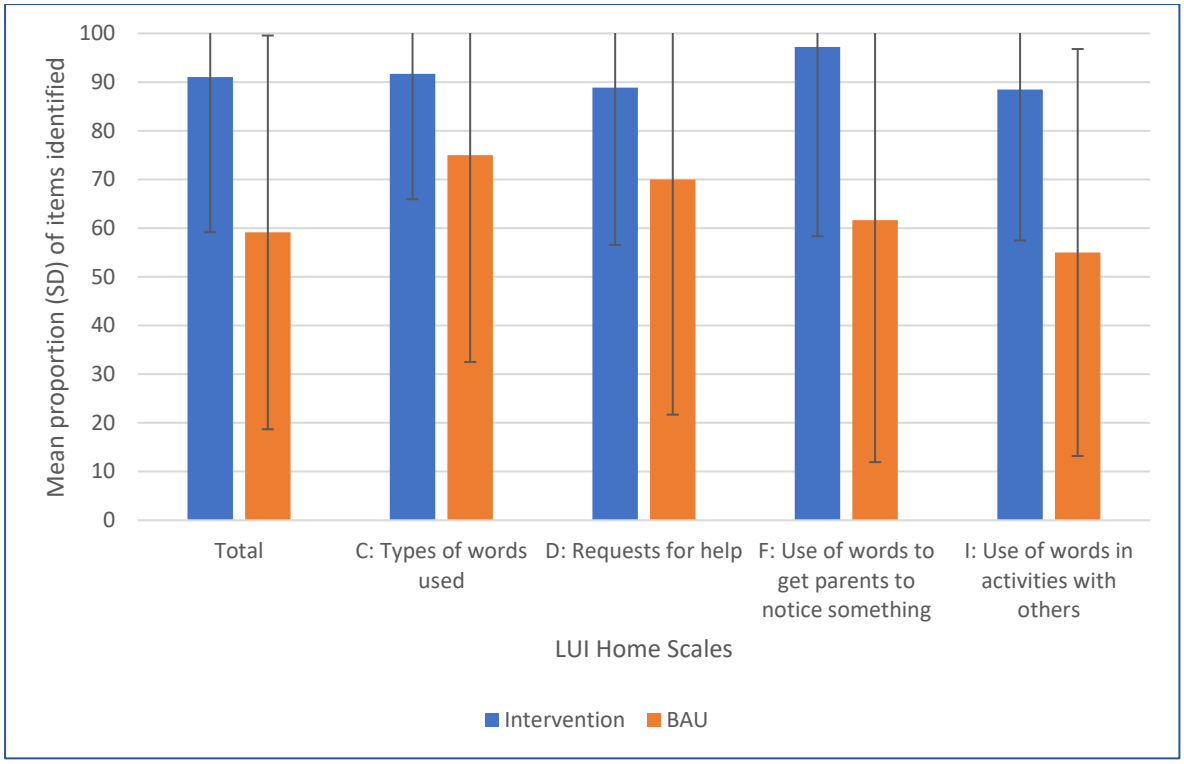

Figure 9: **Parent/carer report of home language ability at follow-up.**

4.5.3.2 Parent/carer Reported Behaviour

There was no significant intervention effect on the Total Difficulties score (p = .900), indicating that the intervention did not have an impact on behaviour. Similar non-significant effects were observed on the individual subscales: Emotional difficulties, Conduct problems, Hyperactivity, Peer difficulties and Prosocial behaviour (all ps > .05).

5. Implementation and process (IPE) evaluation

This section presents findings from the implementation and process evaluation. It considers:

1. How *Talking Time*© was implemented; and how and why implementation varied.
2. How and why the *Talking Time*© programme was adapted by schools and to what extent adaptations were congruent with programme principles.
3. Whether and how schools intended to continue with the programme following the end of the structured intervention period and what challenges they may face.

Of the 18 schools which began the intervention, 14 completed the programme with varying degrees of fidelity. Four dropped out prior to the programme end (the 'leavers'): an intervention attrition rate of 22%. Although direct comparisons cannot be made, it is notable that attrition in this trial was much higher than in the previous 2019-20 trial which began prior to the lockdown. At the time the 2019-20 trial was halted (Week 10), all schools were

still participating. It is possible that the additional pressures of COVID-19 contributed to the higher attrition rate in the current trial.

The IPE was based on surveys, interviews and records from the 14 intervention 'completers', one 'leaver' school, the two trainer-mentors and nine BAU schools. Headlines are shown below, with underpinning data and further detail in the Implementation Annex.

5.1 Implementation fidelity and reach

A high proportion of professional development (PD) components were delivered but significant delays and staff absence due to COVID-19 hampered the schedule:

Delivery of twilight training and in-class mentoring was largely as intended. Schools received 2.72 of 3 training sessions (*SD* = .57) and 2.39 of 3 in-class mentoring sessions (*SD* = 1.04) on average. Video mentoring uptake was more variable with schools receiving 1.61 (*SD* = 1.20) of 3 sessions on average. The 14 'completer' schools received all intended training sessions, 2.86 (*SD* = .36) in-class sessions and 2.07 (*SD* = .92) video sessions on average (Table 4), although adaptations were made in response to staff sickness (Appendix A). The four 'leaver' schools attended some elements of initial preparation for Story Conversations and Word Play but left the programme prior to Hexagons and video mentoring.

Just under half (48%) of professional development sessions were postponed largely due to COVID-19 and/or staff absence. At times, this affected adherence, with some schools running more than a month behind schedule by the end of the programme. Not all schools completed the Hexagons programme.

One further aspect to note is the delivery of the 'orientation' visits prior to the programme start. These were intended to build relationships, prepare staff for the programme, clarify expectations and problem-solve practical barriers prior to the programme beginning. They were planned for late November/early December 2021 to allow a month's lead-in prior to the programme start in early January. However due to COVID-related delays in recruiting schools and completing baseline testing, many (50%) of the orientation visits could not take place until January, particularly in the North East. This compaction of the preparatory period is likely to have influenced school preparedness for the programme (this was confirmed by trainer-mentors in their interviews). Of the four 'leaver' schools, two were oriented in January and one did not complete an orientation at all, following several postponements.

Table 4: Number of professional development sessions received by schools.

No. of sessions	Number of schools receiving *(completer schools in brackets)*		
	Twilight training	**In-class mentoring**	**Video mentoring**
3 sessions	14 *(14)*	12 *(12)*	5 *(5)*
2 sessions	3	3 *(2)*	6 *(6)*
1 session	1	1	2 *(2)*
0 session	0	2	5 *(1)*
Total schools	18 *(14)*	18 *(14)*	18 *(14)*

Trainer-mentors' coverage of PD content was high but hampered by time constraints: Trainer-mentors completed checklists to support standardisation of delivery and monitor fidelity. Records were only available for the Northeast. The proportion of intended content delivered was high: 95% for twilights, 84% for in-class mentoring and 92% for video mentoring. Challenges for content coverage during in-class mentoring across both areas included staff shortages and time restrictions, session interruptions, unavailable pupils (hampering observation) or staff reticence to be observed. Elements missed during twilight training were included in follow-up mentoring. Mentors reported some challenges in providing video mentoring (i.e., technical, practitioner reticence, scheduling).

One Story Conversation twilight session was observed by an external observer in each area to assess fidelity of delivery. These confirmed the high level of fidelity reported by mentors and aligned with mentor records from the same sessions.

5.2 Classroom implementation

5.2.1 What schools delivered

Within most schools, the programme was delivered universally as intended: Eleven completer schools delivered the programme across their nursery class, with the exception of one child with severe needs. Some early misunderstandings about programme in the Northeast meant that three schools offered less than universal coverage.

Overall programme delivery was high for completer schools, but varied across schools: The programme was shortened by one week in response to COVID-related delays (Appendix A). Within the shortened programme schools were expected to offer 15 Story Conversations, eight Word Play and five Hexagons sessions (28 in all). On average, the 14 completer schools delivered almost all intended sessions (M = 97.4%; *SD* = 0.16) reflecting their commitment to the programme in highly challenging circumstances. However, implementation varied,

with some delivering more than intended and others less (range 50.0-118.5%) (Table 5 below, Implementation Annex). Delivery in leaver schools is estimated at < 50%.

Delivery of individual activities was also high for completer schools: 102% for Story Conversations (*SD* = 0.16), 94% for Word Play (*SD* = 0.11) and 90% for Hexagons (*SD* = 0.32). On average, completer schools delivered 15.29 Story Conversation, 7.55 Word Play and 4.5 Hexagon sessions. Delivery in the four leaver schools has been estimated at < 50% (Table 5). Delivery was most variable for Hexagons, which was introduced towards the end of the programme (Table 5), reflecting the impact of COVID-related schedule delays. Some schools which began on time and kept to schedule managed to deliver many more than the planned five Hexagons sessions; while one did not deliver any at all.

Table 5: Proportion of intended Talking Time© sessions delivered.

% of intended sessions delivered	**Number of schools in each category *(completer schools in brackets)***			
	Story Conversations	**Word Play**	**Hexagons**	**Overall**
≥ 100%	11 *(11)*	9 *(9)*	7 *(7)*	7 *(7)*
≥ 80%	2 *(2)*	4 *(4)*	6 *(6)*	6 *(6)*
≥ 50%	1 *(1)*	1 *(1)*	0 *(0)*	1 *(1)*
< 50%[a]	4	4	5 *(1)*	4
Total schools	18 *(14)*	18 *(14)*	18 *(14)*	18 *(14)*

[a] <50% assumed for leaver schools based on feedback from their mentors.

5.2.2 What children received

Average child exposure was high but variable for completer schools. Relatively few children received the full intended quota of sessions, reflecting high levels of child-related COVID-19 absence during the intervention period:

Implementation records were only submitted by completer schools (*n* = 342 children). On average, children in completer schools received 75% (*SD* = 0.33) of intended sessions with exposure highest for Story Conversations (83%) and lowest for Hexagons (65%) (Figure 10). This equates to 12.49 Story Conversations (*SD* = .26), 6 Word Play (*SD* = .29) and 3.22 Hexagons (*SD* = .39) sessions on average per child. However, there was great variability, with some children receiving more than the intended quota and others less (Figure 10). Overall, only 39% of children in completer schools received the 15 Story Conversations sessions intended within the shortened programme, with 33% receiving the intended eight Word Play sessions and 25% the intended five Hexagons sessions. Given the relatively high delivery of *Talking Time*© sessions, this reflects the impact of COVID-related child absences

during the implementation period. The lower rate for Hexagons also reflects the fact that fewer Hexagons sessions were delivered. Child exposure in the four 'leaver' schools is unknown but likely to be considerably lower than exposure in completer schools.

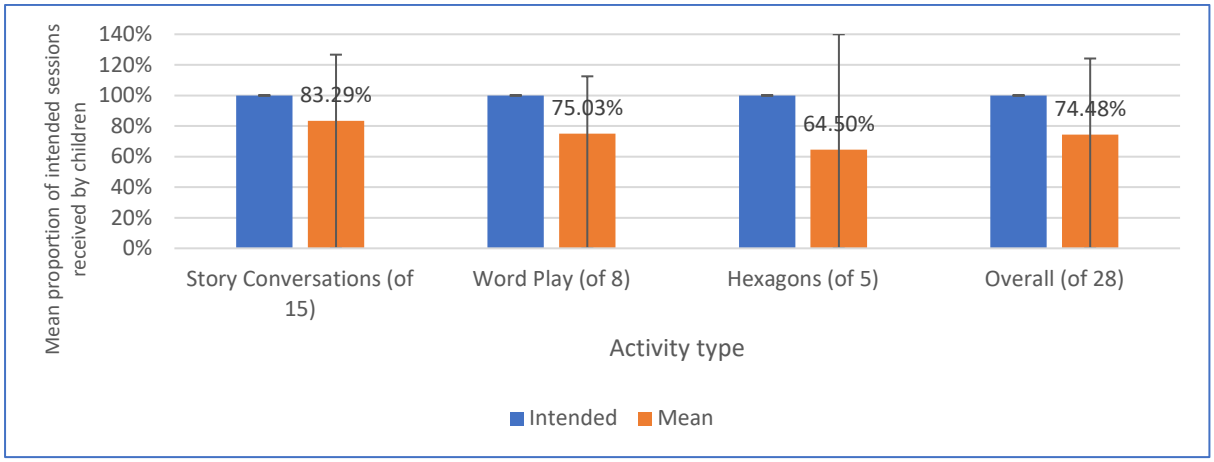

***Figure 10*: Mean proportion and range of sessions received by children in the 14 'completer' schools compared with intended (i.e. 100%).**

Note: These data will reflect a positive skew since no data were available on child exposure for the four 'leaver' schools. Data adjusted for missing weeks. Error bars denote the range, with some extending beyond 100% where children received more than the intended quota.

Schools varied in the degree to which they achieved the intended child exposure:

In 7 of the 18 schools, children received 80% or more of intended sessions (Table 6). In 6 schools, children received 50-80% of intended sessions on average. In 5 schools, children received less than 50% of sessions on average. As already noted, child exposure was lowest for Hexagon sessions, with only 4 schools achieving an average exposure rate of ≥ 80%.

Table 6: Proportion of child exposure to intended Talking Time© sessions.

% of child exposure on average	**Number of schools in each category *(completer schools in brackets)***			
	Story Conversations	**Word Play**	**Hexagons**	**Overall**
≥ 100%	2 *(2)*	0 *(0)*	2 *(2)*	1 *(1)*
≥ 80%	7 *(7)*	7 *(7)*	2 *(2)*	6 *(6)*
≥ 50%	5 *(5)*	6 *(6)*	7 *(7)*	6 *(6)*
< 50%[a]	4	5 *(1)*	7 *(3)*	5 *(1)*
Total schools	18 *(14)*	18 *(14)*	18 *(14)*	18 *(14)*

[a] <50% assumed for leaver schools based on feedback from their mentors.

5.2.3 Adherence and adaptation

Schools generally followed guidance but sometimes faced logistical challenges:

All schools adhered to the programme's small group approach. They also aimed to follow guidance on use of a quiet location, but this was not always achievable, and distractions impacted on child engagement. Adherence to the schedule was sometimes difficult and many schools reported missing classroom delivery of some *Talking Time*© sessions. Reasons for schedule adaptations were generally logistical and commonly COVID-19-related. Staff provided 'catch-up' sessions where possible.

Staff followed manualised planning and developed confidence in adaptation over time:

All schools used the manualised prompts and plans for their sessions. Planning for sessions was brief, never formal and often done in advance. It generally involved the selection of prompts, target vocabulary and any required resources. Reliance on manualised prompts reduced over time and staff gained confidence with planning, adapting questions and tailoring sessions to suit their children's needs. This flexible use of prompts and intentional responsiveness was consistent with programme principles, reflecting the aim to encourage adaptation, ownership and self-sufficiency. Staff also reported increasing confidence in use of language supporting strategies over time.

Practitioners valued the professional learning element of the programme, but time constraints presented a barrier to adherence:

Staff were asked to complete a weekly reflection using the Rehearse, Notice, Analyse and Refine (RNAR) framework, focusing on their use of language-supporting strategies. All staff interviewed reporting reflecting on sessions. This time was valued and seen as a natural part of practice, informing planning of subsequent sessions and improvements in practice. However, unsurprisingly given the wider COVID context, many (n = 25) reported finding it difficult to protect time for professional learning. When asked in the final survey how much time they had been able to set aside to concentrate on their own professional learning, mean ratings were just under 2 (M = 1.92; SD = 0.48) (1=not at all, 2=somewhat, 3=quite a bit, 4=very much). One impact of the reduced time was that reflection happened verbally with colleagues rather than being recorded in staff's Individual Log. Some staff reported finding the paperwork onerous so this may be an aspect to refine in future programme iterations.

5.3 Programme differentiation

***Talking Time*© differed from usual practice in both intervention and BAU schools:**

All completer and BAU schools reported a focus on oracy prior to embarking on *Talking Time*©. However, this generally involved supporting oral language during everyday practice (i.e., continuous provision) combined with targeted withdrawal. Less than half planned small group activities to support oracy.

5.4 Practitioners' perspectives on the programme and influences on implementation

Programme activities and materials were viewed as helpful and supportive:

All activity types were well received by staff. The small group approach was valued, and high levels of child engagement and beneficial outcomes were reported. Story Conversations were particularly valued, with 80% of practitioners identifying this as the activity most benefiting their children. The books were considered a critical resource for promoting engagement, with the wordless books particularly well received. However, staff reported that some books were more valuable for initiating and sustaining conversations than others. The Hexagons resource was generally well regarded, particularly for its simplicity and convenience, although some felt the scenarios were not meaningful for children or found the pictures outdated.

The manualised prompts and plans were rated as helpful for implementation. Staff valued their flexibility and support to adapt practice for individual children. Staff also referred to the usefulness of the timetable and reflective (RNAR) framework.

The professional development (PD) support package fulfilled the needs and expectations of staff. Mentoring was the most highly valued of the PD components:

Practitioners felt well supported, with all but one (n = 25) feeling 'quite' or 'very well' prepared to implement *Talking Time*© in the classroom and the same number feeling 'quite' or 'very well' supported in their professional growth. Trainer-mentors were viewed as friendly, knowledgeable, and dedicated. When asked if the support offered fulfilled their needs and expectations, every practitioner interviewed confirmed this definitively.

All three PD elements were highly rated as being supportive of implementation and professional growth. In-class mentoring was considered to be particularly vital for ensuring implementation quality, promoting professional growth and generating confidence. The opportunities which mentoring provided for first-hand observation, consolidation of initial training and dedicated time for joint analysis of practice were all viewed as powerful support mechanisms.

Schools and mentors reported a wide range of supports and barriers to implementation:

Programme implementation was influenced by a wide range of factors relating to the school, the wider context and to participants themselves (Table 7). Wider environment factors

(staffing, time, COVID-19) presented barriers impacting delivery and child exposure. At child level, high levels of COVID-related child absence and/or children's additional needs presented barriers. Many of the challenges faced by schools were highly inter-related and compounded each other. In particular, the effects of COVID-19 exacerbated existing staffing and time issues. However, schools also identified a range of factors which supported implementation at class level Further detail can be found in the Implementation Annex.

Table 7: Influences on implementation of Talking Time©.

Type of factor	Support	Barrier	Dependent on context
Wider environment		Time Staffing COVID-19	Senior leadership engagement Availability of space
Classroom implementation	Teamworking Timetabling Curriculum congruence		
Implementor (staff)			Preparedness for programme Openness to programme
Child		Child absence Child characteristics/needs	

Staff reported positive impacts on classroom practice and child outcomes, particularly for children with additional language needs:

Staff in all schools interviewed felt their practice had developed because of involvement in *Talking Time©*, including greater expertise working with small groups and in use of language-supporting strategies (particularly in tailoring questions for different children). Staff also felt their practice in sharing stories to support oral language had shifted towards a more child-led approach. Staff in all completer schools reported that children had made progress through participation. A range of perceived benefits were identified. The most commonly noted were vocabulary, increased language confidence and children talking more (e.g., more often, longer sentences, more participation in conversation). The programme was reported to be most beneficial for children with specific language needs, such as children with weak language skills, dual language learners and reluctant communicators.

***Talking Time©* was viewed as feasible for future use:**

Most schools interviewed (n = 12/13) said that the programme was feasible to implement overall. All completer schools were planning to continue with the programme in some form. Staff felt confident to move forwards with the materials provided and also confident

(although less so) to adapt these to match children's needs. A range of recommendations for the future refinement of *Talking Time*© were offered, including in relation to professional learning, programme materials and scheduling.

6. Discussion

We expanded an existing targeted universal oral language programme (*Talking Time*©; Dockrell et al., 2010) to include an evidence-based professional development component and examined its impact in areas of social disadvantage. Impact was assessed at child and practice level and by parental report. Implementation was evaluated.

6.1 Summary of results

6.1.1 Child measures

Social disadvantage is a well-established risk factor for poor oral language development (Law et al., 2018; Nelson et al., 2011). Consistent with extant literature our sample of 3–4-year-old children had, on average, significantly low English language abilities at baseline, confirming the need for a universal oral language intervention in deprived settings. Children's mean scores for language measures increased over time. As predicted, there was no intervention effect on the control measures of non-verbal ability and phonological awareness.

The intervention was partially successful: *Talking Time*© programme had a positive impact on targeted expressive vocabulary. Children in the intervention group made significantly greater progress in their understanding of these words over time compared to children in the business-as-usual (BAU) group. The impact of *Talking Time*© on children's expressive vocabulary (the target object and action words) is noteworthy as, unlike other studies, these words were not formally taught, and staff were not aware there were target vocabulary items. Their use was implicitly embedded in the intervention materials. In the storybooks they appeared either in the text or in the illustrations. In the intervention manual they were embedded – alongside other non-target vocabulary words – within the prompts and plans provided to staff. Our data indicated that vocabulary learning can be supported through adult-child interactions when engaging with story books and toys as well as through formal teaching. Further, the result supports prior research on the benefits of dialogic book reading followed by guided play to reinforce, rehearse and embed vocabulary learning (Dowdall et al., 2020; Toub et al., 2018; Wall et al., 2022). The impact on action words is noteworthy given the relative delay in learning action words and the overuse of general all-purpose verbs by children (Kelly, 1997). Finally our data suggested that language-supporting adult-child

interactions can occur through provision of flexible curricular and professional development materials and an adaptive intervention programme (Bleses, Højen, Dale, et al., 2018). However, no intervention effects were identified for the standardised measures of language, although the effect on Sentence Repetition fell just below significance, an indicator of grammatical competency (Hirsh-Pasek et al., 2015, as cited in Pace et al., 2017) and a key to narrative production. There are a number of potential reasons for the lack of impact on standardised measures. The small sample size meant that analyses were underpowered. The challenges of COVID-19 also meant that the intervention programme did not run as originally intended, despite huge efforts and great commitment from participating schools and the intervention team. The programme as a whole was shortened by one week in response to COVID-related delays. Further challenges included reduced time for school orientation prior to the start of the programme; staffing constraints and absences leading to a high postponement rate for PD and programme schedule delays; constraints on staff capacity to concentrate on their own professional learning; and high levels of child absence limiting child exposure. While many of these challenges are common to any intervention programme, all were substantially exacerbated by COVID-19 during this trial. Due to the low power of the sample we were unable to assess the effect of dosage. However, it is likely that the reduced child exposure, alongside other challenges, will have limited the potential of the programme to impact on child language outcomes. This is particularly true for the Hexagons narrative activity, which was most affected by the delays. Further research under less challenging conditions, with a larger sample size – and also potentially a longer period between baseline and follow-up - may result in a significant intervention effect on the standardised language measures. Future studies should include an analysis of the association between dosage and the outcomes of child language and adult-child interactions.

6.1.2 Parent/carer reports

We predicted that parents/carers in the intervention condition would report higher levels of language and prosocial abilities and fewer behavioural problems at follow-up; and anticipated a small effect in these secondary outcomes by the *Talking Time*© intervention. We also anticipated a low return rate of parent/carer forms in line with previous studies.

6.1.2.1 Language Use

As expected, relatively few parents/carers returned the adapted Language Use Inventory (LUI) at baseline and follow-up. These parents/carers reported that, on average, their children knew the majority of included items in English; but noted fewer examples of children ‘building longer sentences and stories’. When asked about children’s abilities in their home

language, on average children who were DLL knew approximately 70% of the words and phrases on the checklists.

Contrary to predictions, there was no overall change in parent/carer reports of language use over time. As predicted, there was no effect on parent/carer-reported use of English as a result of the intervention. However, there was a significant effect of the intervention on parent/carer report of home language. Due to the limited power of the analyses these findings must be interpreted with caution. Further studies with a larger sample of parent/carer reports would help to examine whether these potential transfer effects from the intervention are valid and reliable.

These findings underscore the importance of engaging with parents/carers in educational research. Findings from parents provided a different perspective to the researcher assessments of child language, particularly in relation to use of children's home language. However, they also highlight the challenges of parent involvement. Not only were the samples from parents small but those most likely to return the LUI forms were monolingual English-speakers with higher levels of education and no concerns about their child's language. This suggests that parent/carer data is missing from the children who are more at risk for poor language development.

6.1.2.2 Socioemotional Behaviour

At baseline, parents/carers reported significantly higher rates of children in the 'High' or 'Very high' categories for behavioural difficulties, compared to population norms. This is in line with previous literature showing elevated social, emotional and mental health difficulties in children from deprived areas (Goodnight et al., 2012; Law et al., 2014; Sim et al., 2013). There was no significant effect of the intervention, time or language status on children's socioemotional behaviour. Again, these findings are derived from a small sample with insufficient power and should be interpreted with caution.

6.1.3 Dual Language Learners

Findings should be interpreted in the light of the fact that our sample had a high proportion of dual language learners (DLL). Recent figures show approximately 21% of primary school pupils in England are children who are DLL (Department for Education, 2022), predominantly found in deprived areas (Strand et al., 2015). These children have been viewed as attending nursery with a 'double disadvantage' in relation to learning English, due to speaking a primary language other than English at home (Hoff, 2013). With just over half of the *Talking Time*© sample speaking another language at home we predicted language status to be a confounding factor in analyses. Children who were DLL had significantly higher scores

on the Non-word Repetition task than their monolingual English-speaking peers. This is in line with previous literature demonstrating a strength in phonological awareness for bilingual children (Balladares et al., 2016). By contrast the monolingual English children had significantly higher scores across all remaining language tasks, although greater growth over time on the Verbal Comprehension and Naming Vocabulary was evident in the children who were DLL. This is a key finding for the sector. While the *Talking Time*© programme had a larger effect on native English-speaking children for the targeted vocabulary items, the children with DLL appeared to benefit more generally from exposure to staff and peers speaking English, highlighting the impact of attending early years settings for children who speak another language at home. Additionally, there is some tentative evidence from parent/carer reports that the intervention may have benefited the home language abilities of children who were DLL. Further work is required to establish whether and how the *Talking Time*© programme can be developed in order to offer equal benefits for all children.

6.1.4 Practice measures

At the practice level, we predicted that the *Talking Time*© programme would enhance nursery staff's interactions with the children in their class as measured within the context of a shared reading activity. The findings support these predictions. The significant increase in the proportion of prompts in the intervention group indicated that intervention practitioners learnt to invite more contributions from children in their groups, which is key given that strategies promoting children's language use can be particularly important for their vocabulary and expressive language development (Justice et al., 2018; Ribot et al., 2018). After inviting contributions from children, the findings also suggest that staff learned techniques for extending the length of these conversations. The number of conversational turns increased significantly more in the intervention group, demonstrating enhanced exchanges which are reported to support children's language development (Cabell et al., 2015; Romeo et al., 2018). One mechanism for achieving this may have been through staff's increased use of prompts; that is, they used prompts successfully to continue conversations as well as initiate them. Finally, the practitioners in the intervention group demonstrated a shift in practice with significantly reduced direct reading from the text whereby they were more able to engage in extended interactions with the children in their groups, in line with the *Talking Time*© principle of co-constructing of stories and conversations with children to enhance oral language.

These improvements in practice appeared to be successful in encouraging greater child participation: a larger growth in the proportion of child utterances was seen in the

intervention group than in the BAU group. In a recent study, the conversational turns which children experienced in their preschool settings were found to be the strongest predictor of their vocabulary skills, with effect sizes comparable to those seen for parent education and coming from English speaking households (Duncan et al., 2022). The fact that improvements in practice were seen despite the challenges reported by staff in protecting time for professional learning and the ongoing COVID situation shows promise for the *Talking Time*© professional development component. It confirms prior studies in showing that bespoke professional development, when well-designed, can successfully improve practice and child language outcomes (e.g., Bleses et al., 2018). It offers initial support for the notion that the contained professional learning context of *Talking Time*© and repeated use of the RNAR framework may offer an efficient context for learning and behaviour change. Future evaluation in less challenging contexts – and assessing a wider range of practice and child outcomes - is needed to further examine the potential of the approach.

6.1.5 Implementation

The first requirement for a successful programme is a demonstrated need for intervention. The baseline data revealed a clear need for targeted universal support for oral language in the early years. Interviews with intervention and BAU schools confirmed that *Talking Time*© differed from usual practice, suggesting potential for the programme to achieve impact. Finally, the enthusiasm with which the programme was taken up and delivered, particularly in the challenging COVID context, demonstrated a desire among schools for targeted universal interventions and effective professional development relating to oral language. The wide-ranging challenges to implementation have already been noted (see 6.1.1) and likely affected session delivery; staff's capacity to improve their knowledge skills and practice; and child exposure. Although many of these challenges were exacerbated by COVID-19, other barriers were identified and findings from the implementation and process evaluation (IPE) provide valuable learning to guide future refinement of the programme. For example, some staff were reluctant to record themselves for the video-mentoring sessions so it may be that this aspect requires further thought and development.

However, despite significant COVID-related challenges, the IPE also strongly confirmed the feasibility and acceptability of the programme. Delivery of the programme among completer schools was high. Over three-quarters of the randomised schools completed the intervention; with all but one of these 'completer' schools delivering 80% or more of the intended sessions. The support for 'principled adaptation' in line with programme principles (Sims et al., 2021) appears to have been successful: staff generally followed manualised

planning, valued the flexible prompts and plans, and developed confidence in adaptation over time. Overall, staff feedback indicated that the programme was largely well-received, acceptable to schools and feasible to implement. Staff felt well supported for implementation and professional growth and reported a wide range of positive impacts for classroom practice and child outcomes, particularly for children with language needs. In line with prior research (Kraft et al., 2018; Sims et al., 2021) the in-class mentoring element was viewed as important for implementation and professional growth.

6.1.5.1 Challenges to implementation

6.2 Next steps

Knowledge mobilisation involves connecting and encouraging early years practitioners and school management to share explicit and tacit knowledge about language learning opportunities and empowering staff to use this knowledge to inform their decision making and activities with the children. To further consider the challenges to the implementation of effective language learning opportunities in the early years we consider the results of the trial in relation to the COM-B model (Michie et al., 2011). This model provides a framework to examine the strengths and limitations of the current study.

Our aim was to enhance practitioner interactions with children to impact on their language skills and behaviour. The COM-B model identifies three factors relevant to behaviour change: capability, opportunity, and motivation (Michie et al., 2011). *Talking Time*© explicitly aims at developing capability within the practitioners and providing evidence about the kinds of opportunities that will enhance language learning within children in their settings. As our evaluation showed, this was both well received by practitioners and, in many contexts, put into practice. By contrast both physical and social opportunities were impacted at the time of the intervention, primarily stemming from the specific challenges of running the trial during the COVID pandemic. Other factors outside of our control hindered the impact of the trial included non-pandemic child and staff illness, staff changeover and children moving schools.

However, there were other physical and social barriers which need to be considered: time. There was a low return rate of audio recordings at baseline and follow-up, with nursery practitioners citing a lack of time and staffing, as well as difficulty finding suitable space to carry out the activities. Furthermore, one of the ‘leaver’ schools reported that the intervention programme did not fit within their current practice. Future iterations of the intervention need to consider what is feasible for nursery settings to provide. Nevertheless, the fact that a twice weekly 15-minute group reading session was not practical is a cause for concern given the

large evidence-base for the effectiveness of regular shared reading and the importance of regular exposure to language learning interactions. Future actions could include knowledge exchange activities to share the benefits of a common elements approach (Clarke, 2022), highlighting the best strategies that can be employed to support oral language development, while at the same time hearing from nursery practitioners about how best to incorporate these aspects into their practice. The more staff are empowered (capability) and can create opportunities the more likely behaviour change is to occur. However, both aspects impact on motivation and high levels of motivation are required to effectively embed language learning interactions.

Time was also a barrier from a different perspective: the relatively short timeframe of the trial meant there was only one time-point for follow-up data collection, so there was no evidence about the longevity of the effects. Equally, due to limited staffing and time, a small handful of codes were applied to the transcripts of adult-child interactions. While the reported findings provide some important insights into how *Talking Time*© impacts on language supporting practice, there is considerable scope to expand on the analysis to provide a deeper insight into its effects. Feedback from practitioners will help to develop the design of the *Talking Time*© programme and further evaluation will provide a more rigorous assessment of the intervention.

6.3 Conclusion

Our project demonstrated the significant need for, and the success of, a universal intervention to improve children's oral language in deprived settings. We captured children's language using direct assessments and parent/carer reports of language abilities in daily life to provide a comprehensive overview of the language abilities of children from deprived areas. At the child level we broadened the typical focus of oral language intervention studies from vocabulary to include measures of comprehension, grammar and narrative and we included dual language learners to provide a more representative and generalisable sample. We also included innovative measures of adult-child interactions in the classroom.

Use of a structured-but-adaptive manualised intervention programme alongside embedded professional development had a positive impact on children's expressive vocabulary and the quality of adult-child interactions during shared book-reading. The significant intervention effect for the bespoke language measure is noteworthy as the vocabulary items were not explicitly taught, highlighting the impact of the strategies that practitioners developed from the intervention. The successful outcomes during very challenging times point to the enthusiasm and commitment of nursery practitioners to provide

a supportive environment for the children under their care, where their language development can flourish. Future studies under less challenging conditions, with a larger sample size and with a longer timeframe may be able to identify further effects for the standardised language measures.

7. References

Altman, D. G., & Bland, J. M. (2005). Treatment allocation by minimisation. *BMJ*, *330*(7495), 843. https://doi.org/10.1136/bmj.330.7495.843

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

Apel, K., & Apel, L. (2011). Identifying Intraindividual Differences in Students' Written Language Abilities. *Topics in Language Disorders*, *31*(1), 54-72. https://doi.org/10.1097/TLD.0b013e31820a22b4

Balladares, J., Marshall, C., & Griffiths, Y. (2016). Socio-economic status affects sentence repetition, but not non-word repetition, in Chilean preschoolers. *First Language*, *36*(3), 338-351.

Bleses, D., Højen, A., Dale, P. S., Justice, L. M., Dybdal, L., Piasta, S., Markussen-Brown, J., Kjærbæk, L., & Haghish, E. (2018). Effective language and literacy instruction: Evaluating the importance of scripting and group size components. *Early Childhood Research Quarterly*, *42*, 256-269.

Bleses, D., Højen, A., Justice, L. M., Dale, P. S., Dybdal, L., Piasta, S. B., Markussen-Brown, J., Clausen, M., & Haghish, E. F. (2018). The Effectiveness of a Large-Scale Language and Preliteracy Intervention: The SPELL Randomized Controlled Trial in Denmark. *Child development*, *89*(4), e342-e363. https://doi.org/https://doi.org/10.1111/cdev.12859

Bleses, D., Jensen, P., Højen, A., Slot, P., & Justice, L. (2021). Implementing toddler interventions at scale: The case of "We learn together". *Early Childhood Research Quarterly*, *57*, 12-26. https://doi.org/https://doi.org/10.1016/j.ecresq.2021.04.008

Burgess, S., Rawal, S., & Taylor, E. S. (2021). Teacher peer observation and student test scores: Evidence from a field experiment in English secondary schools. *Journal of Labor Economics*, *39*(4), 1155-1186.

Buysse, V., Peisner-Feinberg, E., Páez, M., Hammer, C. S., & Knowles, M. (2014). Effects of early education programs and practices on the development and learning of dual language learners: A review of the literature. *Early Childhood Research Quarterly*, *29*(4), 765-785.

Cabell, S. Q., Justice, L. M., McGinty, A. S., DeCoster, J., & Forston, L. D. (2015). Teacher–child conversations in preschool classrooms: Contributions to children's vocabulary development. *Early Childhood Research Quarterly*, *30*, 80-92.

Cabell, S. Q., Zucker, T. A., DeCoster, J., Melo, C., Forston, L., & Hamre, B. (2019). Prekindergarten interactive book reading quality and children's language and literacy development: Classroom organization as a moderator. *Early Education and Development*, *30*(1), 1-18.

Chaparro-Moreno, L. J., Reali, F., & Maldonado-Carreño, C. (2017). Wordless picture books boost preschoolers' language production during shared reading. *Early Childhood Research Quarterly*, *40*, 52-62.

Chen, K., Cheng, Y., Berkout, O., & Lindhiem, O. (2017). Analyzing proportion scores as outcomes for prevention trials: a statistical primer. *Prevention Science*, *18*(3), 312-321.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological assessment*, *6*(4), 284.

Clarke, A., Baker, S., Ghiara, V., Burridge, H., Davie, P., Eberhart, J., Fischer, F., & Jackson, A. (2022). *Common Elements. An innovative approach to improving children's outcomes in early childhood education*.

Cleave, P. L., Becker, S. D., Curran, M. K., Van Horne, A. J. O., & 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.

Coe, R., Aloisi, C., Higgins, S., & Major, L. E. (2014). What makes great teaching? review of the underpinning research.

Cordingley, P., Higgins, S., Greany, T., Buckler, N., Coles-Jordan, D., Crisp, B., Saunders, L., & Coe, R. (2015). Developing great teaching: lessons from the international reviews into effective professional development.

Davies, C., McGillion, M., Rowland, C., & Matthews, D. (2020). Can inferencing be trained in preschoolers using shared book-reading? A randomised controlled trial of parents' inference-eliciting questions on oral inferencing ability. *Journal of child language*, *47*(3), 655-679.

Department for Education, D. (2022). *Schools, Pupils and their Characteristics*.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational researcher*, *38*(3), 181-199.

Dickinson, D. K., & Porche, M. V. (2011). Relation between language experiences in preschool classrooms and children's kindergarten and fourth-grade language and reading abilities. *Child development*, *82*(3), 870-886.

Dietrichson, J., Bøg, M., Filges, T., & Klint Jørgensen, A.-M. (2017). Academic interventions for elementary and middle school students with low socioeconomic status: A systematic review and meta-analysis. *Review of Educational Research*, *87*(2), 243-282.

Dobinson, K. L., & Dockrell, J. E. (2021). Universal strategies for the improvement of expressive language skills in the primary classroom: A systematic review. *First Language*, *41*, 527 - 554.

Dobinson, K. L., Mathers, S. & Dockrell, J. E. (2022). *The quality of universal language support in maintained nursery classrooms in two socially disadvantaged areas of England.* Manuscript submitted for publication.

Dockrell, J., Law, J., & Mathers, S. (2021). Empowering Staff to Enhance Oral Language in the Early Years. *AEA RCT Registry*, *10*, 4379-4372.4370.

Dockrell, J. E., Bakopoulou, I., Law, J., Spencer, S., & Lindsay, G. (2015). Capturing communication supporting classrooms: The development of a tool and feasibility study. *Child Language Teaching and Therapy*, *31*(3), 271-286.

Dockrell, J. E., Forrest, C. L., Law, J., Mathers, S., & Charlton, J. (2022). Screening for language difficulties in disadvantaged populations on entry to early years education: challenges and opportunities. *Frontiers in Pediatrics*, *10*.

Dockrell, J. E., Howell, P., Leung, D., & Fugard, A. J. (2017). Children with speech language and communication needs in England: Challenges for practice. Frontiers in Education,

Dockrell, J. E., Stuart, M., & King, D. (2010). Supporting early oral language skills for English language learners in inner city preschool provision. *Br J Educ Psychol*, *80*(Pt 4), 497-515. https://doi.org/10.1348/000709910x493080

Dowdall, N., Melendez-Torres, G. J., Murray, L., Gardner, F., Hartford, L., & Cooper, P. J. (2020). Shared Picture Book Reading Interventions for Child Language Development: A Systematic Review and Meta-Analysis. *Child development*, *91*(2), e383-e399. https://doi.org/https://doi.org/10.1111/cdev.13225

Druks, J., & Masterson, J. (2000). An action and object naming battery. In: Hove: Psychology Press.

Duncan, R. J., Anderson, K. L., King, Y. A., Finders, J. K., Schmitt, S. A., & Purpura, D. J. (2022). Predictors of preschool language environments and their relations to children's

vocabulary. *Infant and Child Development*, *n/a*(n/a), e2381. https://doi.org/https://doi.org/10.1002/icd.2381

Duncan, R. J., King, Y. A., Finders, J. K., Elicker, J., Schmitt, S. A., & Purpura, D. J. (2020). Prekindergarten classroom language environments and children's vocabulary skills. *Journal of experimental child psychology*, *194*, 104829.

Durham, R. E., Farkas, G., Hammer, C. S., Tomblin, J. B., & Catts, H. W. (2007). Kindergarten oral language skill: A key variable in the intergenerational transmission of socioeconomic status. *Research in Social Stratification and Mobility*, *25*(4), 294-305.

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American journal of community psychology*, *41*(3), 327-350.

Ebbels, S. H., McCartney, E., Slonims, V., Dockrell, J. E., & Norbury, C. F. (2019). Evidence-based pathways to intervention for children with language disorders. *International journal of language & communication disorders*, *54*(1), 3-19.

Elek, C., & Page, J. (2019). Critical features of effective coaching for early childhood educators: A review of empirical research literature. *Professional Development in Education*, *45*(4), 567-585.

Elliot, C., & Smith, P. (2011). British ability scales, BAS3. *London, United Kingdom: GL Assessment*.

Eurostat. (2018). Migrant integration statistics - at risk of poverty and social exclusion. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Migrant_integration_statistics_-_at_risk_of_poverty_and_social_exclusion

Eurostat. (2019). Intergenerational transmission of disadvantages - statistics. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Intergenerational_transmission_of_disadvantages_-_statistics&stable=0#:~:text=(7.6%20%25).-,People%20having%20a%20low%20level%20of%20education%20are%20around%20three,poverty%20levels%20of%20today's%20adults

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.

Fukkink, R. G., & Tavecchio, L. W. (2010). Effects of video interaction guidance on early childhood teachers. *Teaching and Teacher Education*, *26*(8), 1652-1659.

Gardner, H., Froud, K., McClelland, A., & van der Lely, H. K. J. (2006). Development of the Grammar and Phonology Screening (GAPS) test to assess key markers of specific language and literacy difficulties in young children. *41*(5), 513-540. https://doi.org/10.1080/13682820500442644

Girolametto, L., & Weitzman, E. (2002). Responsiveness of child care providers in interactions with toddlers and preschoolers.

Golinkoff, R. M., De Villiers, J. G., Hirsh-Pasek, K., Iglesias, A., Wilson, M. S., Morini, G., & Brezack, N. (2017). *User's manual for the Quick Interactive Language Screener (QUILS): A measure of vocabulary, syntax, and language acquisition skills in young children*. Paul H. Brookes Publishing Company.

Golinkoff, R. M., Hoff, E., Rowe, M. L., Tamis-LeMonda, C. S., & Hirsh-Pasek, K. (2019). Language matters: Denying the existence of the 30-million-word gap has serious consequences. *Child development*, *90*(3), 985-992.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal of Child Psychology and Psychiatry*, *38*(5), 581-586.

Goodnight, J. A., Lahey, B. B., Van Hulle, C. A., Rodgers, J. L., Rathouz, P. J., Waldman, I. D., & D'Onofrio, B. M. (2012). A quasi-experimental analysis of the influence of neighborhood disadvantage on child and adolescent conduct problems. *Journal of abnormal psychology*, *121*(1), 95.

Griffin, T. M., Hemphill, L., Camp, L., & Wolf, D. P. (2004). Oral discourse in the preschool years and later literacy skills. *First Language*, *24*(2), 123-147.

Grolig, L., Cohrdes, C., Tiffin-Richards, S. P., & Schroeder, S. (2020). Narrative dialogic reading with wordless picture books: A cluster-randomized intervention study. *Early Childhood Research Quarterly*, *51*, 191-203.

Harris, J., Golinkoff, R. M., & Hirsh-Pasek, K. (2011). Lessons from the crib for the classroom: How children really learn vocabulary. *Handbook of early literacy research*, *3*, 49-65.

Hassinger-Das, B., Toub, T. S., Hirsh-Pasek, K., & Golinkoff, R. M. (2017). A Matter of Principle: Applying Language Science to the Classroom and Beyond. *Translational issues in psychological science*, *3*(1), 5-18. https://doi.org/10.1037/tps0000085

Hindman, A. H., Wasik, B. A., & Snell, E. K. (2016). Closing the 30 Million Word Gap: Next Steps in Designing Research to Inform Practice. *Child Development Perspectives*, *10*(2), 134-139. https://doi.org/https://doi.org/10.1111/cdep.12177

Hoff, E. (2003). Causes and consequences of SES-related differences in parent-to-child speech.

Hoff, E. (2006). How social contexts support and shape language development. *Developmental review*, *26*(1), 55-88.

Hoff, E. (2010). Context effects on young children's language use: The influence of conversational setting and partner. *First Language*, *30*(3-4), 461-472.

Hoff, E. (2013). Interpreting the early language trajectories of children from low-SES and language minority homes: Implications for closing achievement gaps. *Developmental psychology*, *49*, 4-14. https://doi.org/10.1037/a0027238

Huaqing Qi, C., & Kaiser, A. P. (2003). Behavior problems of preschool children from low-income families: Review of the literature. *Topics in early childhood special education*, *23*(4), 188-216.

Hulme, C., Snowling, M. J., West, G., Lervåg, A., & Melby-Lervåg, M. (2020). Children's language skills can be improved: Lessons from psychological science for educational policy. *Current Directions in Psychological Science*, *29*(4), 372-377.

Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R., & Kerr, K. (2016). Implementation and process evaluation (IPE) for interventions in education settings: An introductory handbook. *Education Endowment Foundation*, *1*.

Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. V. (2010). Sources of variability in children's language growth. *Cognitive psychology*, *61*(4), 343-365.

Joyce, B. R., & Showers, B. (1981). Transfer of training: The contribution of "coaching". *Journal of Education*, *163*(2), 163-172.

Justice, L. M., Jiang, H., & Strasser, K. (2018). Linguistic environment of preschool classrooms: What dimensions support children's language growth? *Early Childhood Research Quarterly*, *42*, 79-92.

Karlsen, J., Hjetland, H. N., Hagtvet, B. E., Braeken, J., & Melby-Lervåg, M. (2021). The concurrent and longitudinal relationship between narrative skills and other language skills in children. *First Language*, *41*(5), 555-572.

Kelly, D. J. (1997). Patterns in verb use by preschoolers with normal language and specific language impairment. *Applied Psycholinguistics*, *18*(2), 199-218. https://doi.org/10.1017/s0142716400009978

König, J. (2013). First comes the theory, then the practice? On the acquisition of general pedagogical knowledge during initial teacher education. *International Journal of Science and Mathematics Education*, *11*(4), 999-1028.

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research*, *88*(4), 547-588.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, 159-174.

Language and Reading Research Consortium (LARRC). (2015). The Dimensionality of Language Ability in Young Children. *Child development*, *86*(6), 1948-1965. https://doi.org/https://doi.org/10.1111/cdev.12450

Language and Reading Research Consortium (LARRC). (2017). Oral language and listening comprehension: Same or different constructs? *Journal of speech, language, and hearing research*, *60*(5), 1273-1284.

Law, J. (2019). The changing perception of communication needs—a litmus Test for the Warnock legacy. Frontiers in Education,

Law, J., Charlton, J., & Asmussen, K. (2017). Child language as a wellbeing indicator.

Law, J., Charlton, J., Dockrell, J., Gascoigne, M., McKean, C., & Theakston, A. (2017). Early Language Development: Needs, provision and intervention for pre-school children from socio-economically disadvantaged backgrounds. London Education Endowment Foundation.

Law, J., Rush, R., King, T., Westrupp, E., & Reilly, S. (2018). Early home activities and oral language skills in middle childhood: A quantile analysis. *Child development*, *89*(1), 295-309.

Law, J., Rush, R., & McBean, K. (2014). The relative roles played by structural and pragmatic language skills in relation to behaviour in a population of primary school children from socially disadvantaged backgrounds. *Emotional and Behavioural Difficulties*, *19*(1), 28-40.

Law, J., Rush, R., Schoon, I., & Parsons, S. (2009). Modeling developmental language difficulties from school entry into adulthood: Literacy, mental health, and employment outcomes.

Law, J., Tulip, J., Stringer, H., Cockerill, M., & Dockrell, J. (2019). Teachers observing classroom communication: An application of the Communicating Supporting Classroom Observation Tool for children aged 4–7 years. *Child Language Teaching and Therapy*, *35*(3), 203-220.

Lendrum, A., & Humphrey, N. (2012). The importance of studying the implementation of interventions in school settings. *Oxford Review of Education*, *38*(5), 635-652.

Levine, D., Pace, A., Luo, R., Hirsh-Pasek, K., Golinkoff, R. M., de Villiers, J., Iglesias, A., & Wilson, M. S. (2020). Evaluating socioeconomic gaps in preschoolers' vocabulary, syntax and language process skills with the Quick Interactive Language Screener (QUILS). *Early Childhood Research Quarterly*, *50*, 114-128.

Markussen-Brown, J., Juhl, C. B., Piasta, S. B., Bleses, D., Højen, A., & Justice, L. M. (2017). The effects of language-and literacy-focused professional development on early educators and children: A best-evidence meta-analysis. *Early Childhood Research Quarterly*, *38*, 97-115.

Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning: A meta-analysis. *Review of Educational Research*, *80*(3), 300-335.

Mathers, S., Malmberg, L.-E., Lindorff, A., & Gardiner, J. (2022). Early educators' knowledge of early language pedagogy: How can it be measured and does it matter for child language outcomes? Oxford: University of Oxford.

McManus, S., Bebbington, P. E., Jenkins, R., & Brugha, T. (2016). *Mental health and wellbeing in England: the adult psychiatric morbidity survey 2014*. NHS digital.

Michie, S., Van Stralen, M. M., & West, R. (2011). The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implementation Science, 6*(1), 1-12.

Mol, S. E., Bus, A. G., & De Jong, M. T. (2009). Interactive book reading in early education: A tool to stimulate print knowledge as well as oral language. *Review of Educational Research, 79*(2), 979-1007.

Nelson, K. E., Welsh, J. A., Trup, E. M. V., & Greenberg, M. T. (2011). Language delays of impoverished preschool children in relation to early academic and emotion recognition skills. *First Language, 31*(2), 164-194.

Neugebauer, S. R., Coyne, M., McCoach, D. B., & Ware, S. (2021). Reframing adherence: Active ingredients and impromptu interactions that support vocabulary implementation effectiveness. *Early Childhood Research Quarterly, 56*, 52-64.

Neuman, S. B., & Dwyer, J. (2011). Developing vocabulary and conceptual knowledge for low-income preschoolers: A design experiment. *Journal of Literacy Research, 43*(2), 103-129.

O'Neill, D. (2009). Language use inventory. *Ontario, Canada: Knowledge in Development, 51*.

OECD. (2018). *The Resilience of Students with an Immigrant Background*. https://doi.org/doi:https://doi.org/10.1787/9789264292093-en

Oracy All-Party Parliamentary Group (APPG). 2021. Speak for Change: Final report and recommendations from the Oracy All-Party Parliamentary Group Inquiry. https://oracy.inparliament.uk/sites/oracy.inparliament.uk/files/2021-04/Oracy_APPG_FinalReport_28_04%20%284%29.pdf

Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of educational psychology, 98*(3), 554.

Pace, A., Luo, R., Hirsh-Pasek, K., & Golinkoff, R. M. (2017). Identifying pathways between socioeconomic status and language development. *Annual Review of Linguistics, 3*, 285-308.

Pellegrino, M., Luisa, M., & Scopesi, A. (1990). Structure and function of baby talk in a day-care centre. *Journal of child language, 17*(1), 101-114. https://doi.org/10.1017/S030500090001312X

Pesco, D., & Gagné, A. (2017). Scaffolding narrative skills: A meta-analysis of instruction in early childhood settings. *Early Education and Development, 28*(7), 773-793.

Petersen, D. B., Staskowski, M., Spencer, T. D., Foster, M. E., & Brough, M. P. (2022). The Effects of a Multitiered System of Language Support on Kindergarten Oral and Written Language: A Large-Scale Randomized Controlled Trial. *Language, Speech, and Hearing Services in Schools, 53*(1), 44-68.

Phillips, B. M., Zhao, Y., & Weekley, M. J. (2018). Teacher language in the preschool classroom: Initial validation of a classroom environment observation tool. *Early Education and Development, 29*(3), 379-397.

Pianta, R., Hamre, B., Downer, J., Burchinal, M., Williford, A., Locasale-Crouch, J., Howes, C., La Paro, K., & Scott-Little, C. (2017). Early childhood professional development: Coaching and coursework effects on indicators of children's school readiness. *Early Education and Development, 28*(8), 956-975.

Piasta, S. B., Justice, L. M., McGinty, A., Mashburn, A., & Slocum, L. (2015). A comprehensive examination of preschool teachers' implementation fidelity when using a supplemental language and literacy curriculum. Child & Youth Care Forum,

Reddy, L. A., Shernoff, E., & Lekwa, A. (2021). A randomized controlled trial of instructional coaching in high-poverty urban schools: Examining teacher practices and student outcomes. *Journal of School Psychology*, *86*, 151-168.

Renfrew, C. (1997). The bus story. *Bicester, England: Winslow*.

Ribot, K. M., Hoff, E., & Burridge, A. (2018). Language use contributes to expressive language growth: Evidence from bilingual children. *Child development*, *89*(3), 929-940.

Røe-Indregård, H., Brinchmann, E. I., Rydland, V., Rowe, M. L., Hagtvet, B. E., & Zambrana, I. M. (2022). Teacher–Child Interactions During Toy Play and Book Sharing. *Early Education and Development*, 1-16. https://doi.org/10.1080/10409289.2022.2142028

Romeo, R. R., Leonard, J. A., Robinson, S. T., West, M. R., Mackey, A. P., Rowe, M. L., & Gabrieli, J. D. (2018). Beyond the 30-million-word gap: Children's conversational exposure is associated with language-related brain function. *Psychological science*, *29*(5), 700-710.

Rowe, M. L. (2022). Environmental influences on early language and literacy development: Social policy and educational implications. In *Advances in Child Development and Behavior* (Vol. 63, pp. 103-127). Elsevier.

Rowe, M. L., & Snow, C. E. (2020). Analyzing input quality along three dimensions: interactive, linguistic, and conceptual. *Journal of child language*, *47*(1), 5-21. https://doi.org/10.1017/S0305000919000655

Shanahan, T., MacArthur, C., Graham, S., & Fitzgerald, J. (2006). Relations among oral language, reading, and writing development. *Handbook of writing research*, 171-183.

Sherin, M. G., & Van Es, E. A. (2009). Effects of video club participation on teachers' professional vision. *Journal of teacher education*, *60*(1), 20-37.

Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard educational review*, *57*(1), 1-23.

Sim, F., O'Dowd, J., Thompson, L., Law, J., Macmillan, S., Affleck, M., Gillberg, C., & Wilson, P. (2013). Language and social/emotional problems identified at a universal developmental assessment at 30 months. *BMC pediatrics*, *13*(1), 1-7.

Sims, S., Fletcher-Wood, H., O'Mara-Eves, A., Cottingham, S., Stansfield, C., Van Herwegen, J., & Anders, J. (2021). What Are the Characteristics of Effective Teacher Professional Development? A Systematic Review and Meta-Analysis. *Education Endowment Foundation*.

Smith, T., Noble, M., Noble, S., Wright, G., McLennan, D., & Plunkett, E. (2015). The English indices of deprivation 2015. *London: Department for Communities and Local Government*, 1-94.

Snow, P. (2014). Oral language competence and the transition to school: Socio-economic and behavioural factors that influence academic and social success. *International journal on school disaffection*, *11*(1).

Systematic Analysis of Language Transcripts (SALT). (2020, October). *Standard Transcription Conventions.* Retrieved January 18, 2023 from https://www.saltsoftware.com/pub/media/wysiwyg/tranaids/StandardConventions.pdf

Strand, S., Malmberg, L., & Hall, J. (2015). English as an Additional Language (EAL) and educational achievement in England: An analysis of the National Pupil Database.

Tomblin, J. B., & Zhang, X. (2006). The dimensionality of language ability in school-age children.

Toub, T. S., Hassinger-Das, B., Nesbitt, K. T., Ilgaz, H., Weisberg, D. S., Hirsh-Pasek, K., Golinkoff, R. M., Nicolopoulou, A., & Dickinson, D. K. (2018). The language of play: Developing preschool vocabulary through play following shared book-reading. *Early Childhood Research Quarterly*, *45*, 1-17.

Turnbull, K. P., Anthony, A. B., Justice, L., & Bowles, R. (2009). Preschoolers' exposure to language stimulation in classrooms serving at-risk children: The contribution of group size and activity context. *Early Education and Development*, *20*(1), 53-79.

van Kleeck, A., Vander Woude, J., & Hammett, L. (2006). Fostering literal and inferential language skills in Head Start preschoolers with language impairment using scripted book-sharing discussions. *Am J Speech Lang Pathol*, *15*(1), 85-95. https://doi.org/10.1044/1058-0360(2006/009)

Wall, D., Foltz, S., Kupfer, A., & Glenberg, A. M. (2022). Embodied Action Scaffolds Dialogic Reading. *Educational Psychology Review*, *34*(1), 401-419. https://doi.org/10.1007/s10648-021-09617-6

Watson, B. (2021). *Employment in the UK: October 2021* (Estimates of employment, unemployment and economic inactivity for the UK., Issue.

Weiland, C., & Yoshikawa, H. (2013). Impacts of a prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. *Child development*, *84*(6), 2112-2130.

Weiler, B. K., & Decker, A. L. (2022). The Impact of SES on Language Domain in Kindergartners' Quick Interactive Language Screener (QUILS) Performance. *Communication Disorders Quarterly*, *43*(2), 133-138.

Wojcik, E. H. (2018). The development of lexical–semantic networks in infants and toddlers. *Child Development Perspectives*, *12*(1), 34-38.

Wright, H., Carr, D., Wiese, J., Stokes, L., Runge, J., Dorsett, R., & Heal, J. (2020). Using Research Tools to Improve Language in the Early Years: Evaluation report and executive summary. Education Endowment Foundation: London, UK.

Zaslow, M., Tout, K., Halle, T., Whittaker, J. V., & Lavelle, B. (2010). Toward the Identification of Features of Effective Professional Development for Early Childhood Educators. Literature Review. *Office of Planning, Evaluation and Policy Development, US Department of Education*.

Zucker, T. A., Cabell, S. Q., Justice, L. M., Pentimonti, J. M., & Kaderavek, J. N. (2013). The role of frequent, interactive prekindergarten shared reading in the longitudinal development of language and literacy skills. *Developmental psychology*, *49*(8), 1425.