A cluster randomised controlled trial comparing the efficacy of pre-school language interventions - Building Early Sentences Therapy and an Adapted Derbyshire Language Scheme

Short title: comparing the efficacy of preschool language interventions

Cristina McKean[a, f], Christine Jack[a], Sean Pert[b], Carolyn Letts[a], Helen Stringer[a], Mark Masidlover[c], Anastasia Trebacz[d], Robert Rush[e], Emily Armstrong[a], Kate Conn[a], Jenny Sandham[a], Elaine Ashton[a], Naomi Rose[a]

a) Newcastle University, Newcastle upon Tyne, UK b) University of Manchester, Manchester UK c) Derbyshire language Scheme, Derbyshire, UK d) Sunderland University, Sunderland, UK e) Finn Coral Statistical Consultancy, UK f) Department of Education, University of Oxford

Acknowledgements

This research was funded by the Heather van der Lely Foundation. We thank the foundation for this funding and for their flexible and supportive approach during the COVID pandemic. Funding for the development of Building Early Sentences Therapy (BEST) was provided by Newcastle University. We thank the many Early Years Practitioners, Schools, Parents and Children for their participation in this study. Their enthusiasm, commitment and support for data collection and intervention delivery before and during the COVID pandemic was truly humbling. The first author would like to acknowledge the enormous professionalism, flexibility and resilience of the RA team (CJ, EA, KC, JS, NR, EA) to continue to deliver the interventions during the pandemic and their commitment to providing high quality and safe intervention to the children in the study at a time when the needs were so high. Thank you.

Conflict of interest statement

BEST materials are freely available for download from https://research.ncl.ac.uk/lively/interventions/best/. No profits are made by the authors. The Adapted Derbyshire Language Scheme (A-DLS) Manual is available on the DLS website. MM receives royalties from the sale of DLS materials purchased on the site.

Data availability statement

Data are deposited in the Newcastle University data repository and are openly available for use by other researchers https://data.ncl.ac.uk/

McKean, Cristina; Jack, Christine; Pert, Sean; Letts, Carolyn; Stringer, Helen; Masidlover, Mark; et al. (2025). Language Intervention in the Early Years Randomised Controlled Trial. Newcastle University. Dataset.

https://data.ncl.ac.uk/articles/dataset/Language_Intervention_in_the_Early_Years_Randomised_Controlled_Trial/28270583

Correspondence to Professor Cristina McKean, Department of Education, 15 Norham Gardens, University of Oxford, Oxford, OX2 6PY Phone: +44 (0) 1865 274024 email: cristina.mckean@oxford.ox.ac.uk

Cristina McKean ORCID: 0000-0001-9058-9813

Helen Stringer ORCID: 0000-0002-7470-2166

Anastasia Trebacz ORCID: 0000-0002-0528-6082

Abstract

Background Children's language abilities set the stage for their education, psychosocial development and life chances across the life course. This study compares the efficacy of two preschool language interventions delivered with low dosages in Early Years Settings (EYS): Building Early Sentences Therapy (BEST) and an Adapted Derbyshire Language Scheme (A-DLS). The former is informed by usage-based linguistic theory, and the latter by typical language developmental patterns. **Methods:** We conducted a preregistered cluster randomised controlled trial in 20 EYSs randomised to receive BEST or A-DLS (https://doi.org/10.1186/ISRCTN10974028). Children aged 3;05–4;05, who were monolingual, with comprehension and/or production scores ≤16th centile (New Reynell Developmental Language Scales (NRDLS)) and no sensorineural hearing impairment, severe visual impairment or learning disability were eligible. 102 children received the intervention. Speech and Language Therapists delivered interventions with high fidelity in 15-minute group sessions twice weekly for eight weeks. Baseline (T1), outcome (T2), and follow-up (T3) measures were completed blind to intervention arm**.** Outcomes were NRDLS comprehension and production standard scores (SS), measures of language structures targeted in the interventions and communicative participation (FOCUS-34). **Results:** Both interventions were associated with significant change from T1 to T2 and T1 to T3 in all outcomes. There were no differences between interventions in gains in NRDLS comprehension SS at T2 or T3. BEST produced greater gains in NRDLS production SS between T1-T2 (d = .40) and T1-T3 (d = .55) and in BEST targeted sentences (d = .77). Children receiving BEST made significantly more progress after intervention (T2-T3) in both comprehension and production. Both interventions were associated with large, clinically significant changes in communicative participation as measured by teacher report (FOCUS-34). **Conclusions:** A low-dosage intervention can produce language gains with moderate to large effects. The accelerated progress after the BEST intervention underscores the significant potential of interventions designed with reference to usage-based theory, which precisely manipulate language exposure to promote the specific cognitive

mechanisms hypothesised to promote language learning. **Keywords:** language; preschool; RCT; intervention; usage-based

Introduction

Children's language abilities set the stage for their education, psychosocial development and life chances across the life course. Children with low language at school entry have substantially increased risks of difficulties with literacy, educational attainment, mental health, quality of life, social inclusion and employment (Hulme et al., 2015; Law et al., 2009; Le et al., 2021; Schoon et al., 2009, 2010; Tomblin, 2014; Törnqvist et al., 2009). The social gradient in language abilities due to the effects of poverty and wider family socio-economic circumstances has long been recognised (Reilly & McKean, 2023; Reilly et al., 2014). It has been brought into yet sharper focus by the COVID pandemic with effects of social restrictions disproportionately affecting the language of socially disadvantaged children (Tracey et al., 2022) bringing extreme pressures to bear on a depleted early years workforce (Axford et al., 2015; Eadie et al., 2021; Early Years Alliance, 2021).

Several preschool language interventions have proven efficacy (Bleses et al., 2018; Frizelle, Mullane, et al., 2021; Law & Charlton, 2022; Law et al., 2017; West et al., 2024) with small to moderate effect sizes. However many do not fit available resources or service delivery models, making implementation difficult and inequitable (Greenwood et al., 2020; McKean & Reilly, 2023; Snowling et al., 2022). Children's services need access to a range of interventions with proven efficacy to choose approaches that best fit the needs of the populations they serve, align with the constraints of service provision, and bring lasting benefits to children.

This study compares the efficacy of two preschool language interventions delivered with low dosages in Early Years Settings (EYS) (Frizelle, Tolonen, et al., 2021b). Head-to-head comparisons of interventions are rare but provide valuable practical and theoretical insights (Frizelle, Tolonen, et al., 2021a, 2021b). Comparisons of effective interventions enable informed choices regarding which works best for a given child, context, family preference or outcome. Comparing interventions

delivered with the same dosage, delivery context, level of treatment fidelity and similar resources, tests whether it is the specific learning mechanisms/active ingredients exploited by the interventions which promote change or simply 'therapy general' effects (Frizelle & McKean, 2022). This study compares the efficacy of two interventions: Building Early Sentences Therapy (BEST) (McKean et al., 2013) and an adaptation of the Derbyshire Language Scheme (DLS) (Knowles & Masidlover, 1982). Both interventions aim to develop children's use and understanding of simple sentences with 2, 3 and 4 clauses. In a quasi-experimental pilot study, Trebacz et al. (2023) found that BEST produced greater standard scores gains in expressive language than a treatment-as-usual control but not comprehension. In an RCT Broomfield and Dodd (2011) demonstrated that DLS was associated with improvements in comprehension but not production when compared to a wait-list control. Comparison between two active interventions is clearly a more stringent research design than comparisons with waitlist controls or treatment as usual, bringing smaller effect sizes but also greater confidence that any differences found can be attributable to the intervention.

BEST is based on usage-based linguistic theory (Tomasello, 2000) and systematically manipulates the nature and quantity of the language a child hears to promote the development of abstract representations of predicate-argument structures (PAS) and hence enable the flexible use of a range of sentence structures (McKean et al., 2013). By promoting abstract representations of PAS, the authors hypothesise that BEST can accelerate future language learning (Langacker, 2000) through the memory and processing advantages which abstract knowledge affords (for detailed theoretical background see McKean et al., 2013; Trebacz et al., 2023).

Usage-based or constructivist theories posit that the adult end state of language acquisition consists of an inventory of constructions linked to the pragmatic and semantic functions they communicate, rather than a set of grammatical 'rules' (Croft & Cruse, 2004). These constructions vary along a continuum of abstractness with respect to the lexical items which can be placed into them; constructions range from the highly concrete and inflexible (e.g. 'How do you do?') to the highly

abstract, and flexible (e.g. NOUN1 + VERB + NOUN2 – meaning NOUN1 acts on NOUN2 and NOUN2 is affected), with other constructions falling somewhere in between (e.g. X wouldn't Y let alone Z).

Tomasello described a usage-based, constructivist account of language acquisition from words to adult 'grammar' (Tomasello, 2000, 2003). Once multi-word utterances begin to be used, language constructions are posited to proceed through five phases: 1) frozen phrases; 2) lexically specific constructions; 3) abstract constructions; 4) paradigmatic categories and 5) retreat from over-generalisation. BEST aims to support preschool children to develop their knowledge and representations of 2, 3 and 4-clause sentences and move through the first three stages from frozen phrases to item-based constructions to abstract representations.

Tomasello's account also describes the cognitive mechanisms brought to bear on the learning process which allow children to move from one stage to the next. BEST manipulates the language input and learning context to support the child's use of these cognitive mechanisms. BEST exaggerates the qualities of the input and provides additional cues to make these cognitive mechanisms more available to children with and at risk of language difficulties. BEST is designed to exaggerate the features which promote intention reading, cultural learning, categorisation, schematisation, and analogy, and promote mapping and retention, thus supporting the development of abstract representations. Active ingredients manipulated in BEST include the use of joint action routines with turn-taking; modelling of actions with toys to support mapping of meaning and of argument structure roles; massed modelling of sentences with systematic variation of nouns around verbs; alignment of sentence models with the same predicate-argument structures but differing verbs; signing both content words, to support mapping, and morphology to draw attention to the morphological frame. The stages of multiword development, relevant cognitive mechanisms and relevant 'active ingredients used in BEST are summarised in Appendix 1.

The DLS, widely used in the UK (Knowles & Masidlover, 1982; Roulstone et al., 2012) is based on research describing the stages of typical language development (Bloom & Lahey, 1978; Brown,

1976). DLS provides a structured syllabus of activities that is individualised to the child's language level, aiming to improve both comprehension and expression. Children are supported to understand and use sentences of increasing length and complexity through play-based activities where the number of 'information-carrying words' a child is asked to understand or use, gradually increases, beginning with their current 'word level' and building incrementally. A structured language teaching approach is taken where a game is created to practice understanding and elicit expressively the specific language structure being targeted. The language used should look and sound as natural as possible, and the games are created so that there is a real pragmatic motivation for the child to engage in the activities and the communication which is central to the game. Children progress through 'word levels' (WL) indicated by the number of information-carrying words in the target sentences (e.g. 1WL – objects; actions; 2WL – object + place; object action; 3WL person + action + object; object + place (including adjective); 4WL – person + action + place (including adjective)). Joint action routines are created within structured activities which constrain the language used and the language required to be understood. Role reversal is used such that children take turns with the adults and other children to follow instructions and take the role of teacher to provide instructions or descriptions. A range of prompting and support if children make errors and/or to promote progression are detailed in the program, including 'bridging' where tasks are made easier by manipulating the context to reduce the WL, cloze procedures, recasting, and error correction.

Based on previous research we hypothesise that BEST and an adapted DLS (A-DLS) will be associated with positive change, with greater gains in children's production from BEST and in comprehension from DLS. Based on underpinning theory and due to the hypothesised promotion of abstract representations allowing knowledge to be generalised, we hypothesise that BEST will bring greater benefits in non-targeted structures, and accelerated progress after the intervention.

Research Questions

1. Which intervention brings greater gains in language production and comprehension?

2. Do interventions differ in the degree to which benefits transfer to non-targeted language structures and/or communicative participation?
3. Do interventions differ in the degree to which language abilities continue to improve post-intervention?

Methods

This preregistered cluster randomised controlled trial took place in three local authorities (LAs) in England between January 2020 and June 2022 (ISRCTN10974028) and is reported with reference to CONSORT guidance (Campbell et al., 2012) (Appendix 2). Twenty Early Years Settings (EYS) were allocated to receive either BEST or A-DLS in two waves to avoid contamination within an EYS and enable group delivery. A simple power calculation using Cohen's power tables (Cohen, 1988) was completed. At 80% power, two-tailed α of .05 and an estimated effect size of d = .5, (derived from Hagen et al. (2017) the most similar recent trial), the sample required was sixty-five children in each arm. A target of seventy-two in each treatment arm was set to allow for the 6% attrition found in a study pilot which was conducted in three settings in areas of social disadvantage through student dissertations at Newcastle University. The aims of the pilot were to determine the most appropriate outcome measures, levels of need, recruitment and retention rates and acceptability of study processes to parents/caregivers and settings. As the study coincided with the COVID pandemic some modifications were needed, details of which are provided in the registered protocol (see ISRCTN10974028). The most significant change was the removal of a Treatment as Usual (TAU) arm. The team and participating schools felt it was unethical to assess children's language without offering additional support and intervention at this time when children's language, communication and social-emotional wellbeing were at such high levels of risk. The sample in each treatment arm was slightly lower than originally planned due to the loss of time when the UK was in full lockdown, reducing the number of waves of active data collection from three to two.

Treatment was delivered by Research Assistants (RAs), who were qualified speech and language therapists (SLTs), to 102 preschool children twice a week for eight weeks: ten EYSs in each of two waves. Newcastle University's ethics committee gave ethical approval. Parents/carers, headteachers, and EYS staff provided fully informed consent.

Recruitment of EYS

LA Early Years advisers and/or Speech and Language Therapy Managers were approached to act as gatekeepers and asked to invite EYSs they identified as having high levels of need to an information event (i.e. where there were known to be high proportions of children not meeting the UK statutory assessment Early Years Foundation Stage Profile (EYFSP) expected levels for Communication and Language ; and/or had high levels of referrals to and requests for advice from Speech and Language Therapy services). Thirty-six EYSs completed an expression of interest to be considered for the study. EYSs with the proportion of bilingual children higher than the average in England (20%) were excluded (n = 8) and invited to participate in a study for non-English delivery of BEST. Twenty-four from the remaining 28 were chosen at random for participation by a statistician external to the study as this was the maximum number of settings where intervention could be delivered within the capacity of the RA team. After the pause in the study due to the COVID pandemic, seven EYSs withdrew and a further four were invited to join. The resulting twenty-one EYSs were randomised to receive BEST or A-DLS.

Randomisation

Randomisation of EYSs to one of two intervention arms was conducted by a statistician not involved with the study. Ten settings were randomised for wave 1 participation and 11 for wave 2. To enable the delivery of the intervention to the maximum number of children, nurseries were grouped into geographical clusters.. Randomization applied the minimization method (Altman & Bland, 2005), stratifying by geographic cluster and social disadvantage (High/Low). High and Low social disadvantage was assigned using a median split in the proportion of pupils in each school eligible for

Pupil Premium: a UK government school subsidy provided for children meeting criteria of social disadvantage (www.gov.uk/government/publications/pupil-premium/pupil-premium). Minimization aims to balance these factors across the treatment arms (Table 1). This technique randomly chooses participants from the available pool (in this case EYSs) and then assigns them to intervention arms in turn in a manner which best maintains the balance between the groups with respect to geography and social disadvantage. The balancing also included a random aspect, of 75%, assignation to that arm which would minimise the difference. One EYS randomised to BEST in wave 2 withdrew from the project.

Recruitment of children

EYS staff were asked to identify children who met the following criteria: age 3;05–4;05; monolingual speaker of English or English as primary language; language development below age-related expectations based on practitioner judgement and consideration of EYFSP guidance; able to participate in small group learning; no sensorineural hearing impairment, severe visual impairment or diagnosed learning disability. Staff approached parents/carers of children they judged met study eligibility criteria, sharing information and consent forms. Parents/carers were told which arm of the study their child's setting had been allocated to before they signed up for the study. Once consent was obtained, children were blind assessed by RAs (T0) and only included in the interventions if they met the following additional inclusion criteria: demonstrated symbolic play, triadic attention and sufficient attention and turn-taking ability to participate in small group activities and scored at or below the 16th centile for production and/or comprehension on the New Reynell Language Development Scales (NRDLS) (Edwards et al., 2011).

Measures

Children were assessed by RAs blind to treatment arm allocation for eligibility (T0), before the intervention (T1), immediately after the intervention, (T2) and at follow-up (T3 approximately 9 weeks after T2). The average gap between T1 and the start of intervention for most EYSs (n=15) was

3.8 weeks (SD = 1.6; range: 1 – 7). However, for 5 EYSs (3 BEST; 2 A-DLS), there was a gap of ~11 weeks due to unavoidable staffing changes during the study.

The outcomes were oral language development and communicative participation. At T1, T2, and T3, children were tested on standardised measures of receptive and expressive language and their knowledge of language structures targeted in the interventions, and teachers and parents asked to report on the children's communicative participation.

NRDLS is a standardised, normed reliable and valid omnibus language assessment that measures young children's comprehension and production abilities yielding standard scores (SS) (Edwards et al., 2011).

BEST Assessment is a probe designed to monitor progress. Children describe sixteen images representing the verbs targeted in the intervention. The images differ across the three assessment time points, assessing the same structures using different noun vocabulary to reduce practice effects. The child's response to the pictures is transcribed then scored with respect to the proportion of content words and morphology used correctly, with a maximum possible raw score of 115 (McKean et al., 2013).

Adapted Derbyshire Language Scheme Rapid Screening Test The Derbyshire Language Scheme includes a Rapid Screening Test of children's ability to follow instructions to manipulate toys which contain 1, 2 and 3 information carrying words (ICW) (e.g. show me the cup; put the spoon in the cup; put the pencil under the box) and commands containing 'and' in lists and sequences of instructions (Knowles & Masidlover, 1982). The child's word level, which is the highest number of ICWs with at least half the test items correct, indicates the starting point for intervention for those receiving A-DLS. Additional instructions containing 4 ICWs were added (e.g., put the big key on Teddy's plate), combined with scores from 44 items from NRDLS, which assess comprehension of 2, 3, and 4 ICWs, to gain a sensitive measure of change due to intervention. A raw score from the total of 69 items was derived as the outcome.

Functional Outcomes in Children Under Six - FOCUS-34 (Thomas-Stonell et al., 2012) was completed by parents and teachers to provide a measure of communicative participation in different areas of their life (home, nursery, playing with friends). FOCUS-34 has two sections, the first measures how well the child communicates in their daily life, whilst the second measures how much help the child needs to do certain things. The parent or teacher uses a 7-point scale ranging from 'not at all like my child' to 'exactly like my child'. The FOCUS-34 does not generate *standard scores* but enables the evaluation of change over time and the identification of meaningful clinical change. This is defined as changes in the child's function that are considered to be important to both SLT and parents/carers. A difference in scores pre- and post-intervention >16 is classified as a significant clinical change.

Vineland Adaptive Behaviour Scales - Vineland-3 were completed by the child's teacher at T0, to characterise children's nonverbal and broader developmental profiles (Sparrow et al., 2016). This questionnaire measures personal and social skills needed for everyday life. Domains include communication, daily living skills, socialisation, and maladaptive behaviours and combine to provide an Adaptive Behaviour Composite.

Interventions

Building Early Sentences Therapy is usually delivered in small groups (3-6 children) but can be delivered one-to-one. Sixteen sessions of approximately fifteen minutes duration are delivered twice weekly for 8 weeks. BEST aims to improve children's use and understanding of two, three and four-clause sentences (i.e. 2: The girl is jumping; 3: the boy is eating a banana; 4: the baby is putting the cup on the table) (See Appendix 3). The intervention exposes children to models of the target sentences in a controlled way, involving both massed and distributed exposure, with controlled variation within the target sentences and controlled contrast between the sentences heard, all presented within a joint action routine. Within each session, children are taken through a two-phase process three times. In phase 1 (Input with variation) the child hears Verb 1 (e.g. eat) of the target predicate-argument structure (e.g. Agent + Action + Patient) used 3-6 times with a frame held

constant and one slot varied (e.g. The man is eating an *apple*, the man is eating an *orange*, the man is eating a *banana*). Whilst hearing the input the child sees the actions being completed by the adult with miniature toys. Sign is used alongside speech signalling content (Makaton (Walker, 1987)) and morphology (Paget Gorman Signed Speech - (Rowe, 1981)). In phase 2 (Output with variation and contrast), the child watches the adult act out an event with the same PAS as phase 1 but with a contrasting verb, and the child is encouraged to describe what they see and the adult recasts their attempt verbally and with sign (again signalling content (Makaton (Walker, 1987)) and morphology (Paget Gorman Signed Speech - (Rowe, 1981)). The child is then allowed to act out the event with the toys while the adult again provides a model of the target utterance. This is repeated a number of times, again with a frame held constant and one slot varied. Following each session parents are given a homework booklet containing pictures of the verbs targeted in the session with a range of agents and patients. Parents/carers are encouraged to describe the pictures and so provide repeated input of the target sentences. The child is not expected to repeat or imitate these sentences but is praised and rewarded if they do so spontaneously. A video explaining and demonstrating the homework was made available and parents were texted as reminder after each session with a link to the video. The intervention is described in greater detail in Trebacz et al. (2023) and the manual, intervention and homework resources (McKean et al., 2013) are available from https://research.ncl.ac.uk/lively/interventions/best/

Adapted Derbyshire Language Scheme. The Derbyshire Language Scheme (Knowles & Masidlover, 1982) is a flexible syllabus of structured play-based activities which is individualised to a child's profile of language skills considering both their comprehension and production abilities. The content of and progression through the syllabus is based on typical language development and the work of Bloom (Bloom & Lahey, 1978; Bloom et al., 1975) and Brown (Brown, 1976; Brown et al., 1981). The scheme includes assessment materials to determine the child's level of abilities, their intervention starting point, and to monitor their progress. Children's progression is individualised through stages

increasing their understanding and use of sentences with 1, 2, 3 and 4, information-carrying words. Play-based activities designed to be meaningful to preschool children are used where children are encouraged to verbally direct other children or the teacher. We created an adapted version of DLS (A-DLS) which could be delivered with high treatment fidelity and reliability in a research context and which matched BEST as closely as possible in terms of dosage and delivery whilst retaining DLS key principles and characteristics. Masidlover, one of the original creators of DLS, created new DLS materials for each activity and provided detailed feedback and advice in the development of the manual and approach. The manual is available at https://www.derbyshire-language-scheme.co.uk/AdaptedDLSManual.pdf . The activities and sessions were piloted and refined before the trial commenced. A-DLS was delivered in small groups with children at either 1-2, 2-3 or 3-4 word level. Homework packs for each activity were developed and provided together with guidance videos for parents (https://research.ncl.ac.uk/lively/interventions/dls/dls/). Parents were sent a text message reminding about the homework, which specific pack to choose to reinforce the work done in the session and linked to the relevant video where explanations and models were provided. A-DLS differs from traditional DLS in that children move more rapidly through the range of DLS target sentences and is less individualised in terms of progression and modification of resources. More details of these differences are available on the study website (https://research.ncl.ac.uk/lively/interventions/dls/theadapteddls/).

Treatment Fidelity

Detailed manuals, scripts for each session, and recording forms were developed for both interventions and a standard set of toy resources and homework materials created. Observational rating scales were used to assess RAs fidelity to the intervention (see Baker et al., 2023 for further detail). Prior to intervention delivery in the trial RAs were trained in the interventions by CM and SP; this included video recording their delivery of interventions, reflecting on their fidelity and receiving feedback from SP and CM using that rating scale. Over the course of the study, fidelity was checked

in the same way for all intervention groups in week 2 of intervention delivery. Due to video recording failure in a small number of sessions 98.02% of groups were rated. It must be noted that some aspects of delivery were affected by the COVID pandemic. All soft toys were necessarily replaced by toys which could be disinfected and resources were sanitised after each session. The RAs wore gloves, aprons and visors designed for use in paediatric care (i.e. decorated with animals). Children were shown videos to help them to prepare for how the RAs would be dressed. RAs used their skills and experience as SLTs to ensure the children were comfortable and to build rapport.

Analysis

All analyses were completed using ANCOVAs to compare outcomes between groups at T2 and T3 adjusting for any group differences which were present in relevant outcomes at T1. Bootstrapping (1000 samples), with replacement, was used to provide Standard Errors and 95% Confidence Intervals for the respective regression coefficients. This approach can be used when there are small samples. It repeats the analysis with randomly drawn new samples from the available data set and in so doing creates more robust estimates with associated confidence intervals. Given the small sample size bootstrapping was preferred over robust standard errors (Mansournia et al., 2020). Multi-level modelling was considered to account for clustering within schools but rejected due to the small cluster sizes within schools (range 1-10) and small total sample size (Maas & Hox, 2005; McNeish, 2014). These analyses were repeated for NRDLS Production and Comprehension Standard Score covarying Wave and Delayed Intervention to assess for potential confounds linked to the pandemic-related differing experiences of the children receiving interventions in wave 1 and wave 2, and for delayed intervention which occurred in 5 schools. Within-group analyses over different time periods were conducted with paired t-tests and checked with repeated measures ANOVA.

Results

Participants

One hundred and seventy-eight children were put forward by EY practitioners as children who were not meeting age related expectations according to the UK Early Years Foundation Stage Profile and curriculum. Parental consent was given for 144 children who were then assessed for eligibility according to inclusion and exclusion criteria. One hundred and three met those criteria, one left the study (Figure 1).

One hundred and two children across twenty schools received intervention: forty-four receiving BEST and fifty-seven receiving A-DLS. School and participant demographic characteristics are presented in Table 1. There were no significant differences between intervention groups with respect to EYS SES, numbers per setting, children's gender, age and non-verbal abilities. There was a significant difference in group sizes across interventions, with smaller groups in the DLS arm due to the greater individualisation of the intervention tasks. There was no significant difference between groups at baseline for NRDLS production scores (BEST M = 75.5; A-DLS M = 75.1) but there was a significant difference for NRDLS comprehension scores, with higher scores in the BEST group (BEST M = 83.2; A-DLS M = 77.4). Analyses reported below adjust for these baseline differences. Using the Indices of Deprivation Affecting Children (IDACI) (Ministry of Housing Communities and Local Government, 2019) as a measure of social disadvantage suggests the participants are relatively socially disadvantaged (66% in quintile 1 (Q1 -most disadvantaged 20% in England), 12% in Q2; 7% Q3; 3% Q4; 12% Q5). The median IDACI decile for the children in the A-DLS arm was lower than the BEST arm however this difference was not statistically significant.

----------Table 1. Participant Characteristics----------

----------Figure 1. CONSORT diagram ----------

Intervention delivery

RAs delivered on average 15.86/16 sessions (SD = 0.34, range: 15–16). These were similar for BEST (M = 15.95/16, SD = 0.21. range: 15–16) and A-DLS (M = 15.79/16, SD = 0.41. range: 15–16). Missed sessions were in the main due to COVID pandemic restrictions disrupting EYS provision. Some children did not attend all offered sessions due to EYS absence. Overall children received an average of 13.44/16 sessions (SD = 2.08 range: 6–16). These were similar for BEST (M = 13.34/16, SD = 2.29. range: 7–16) and A-DLS (M = 13.51/16, SD = 1.91/16. range: 6–16). Treatment fidelity was high with average percentages on the rating scales (where 100% represents perfect fidelity) 97.82 (SD = 3.84). BEST average fidelity was 99.00 (SD = 1.94) and A-DLS fidelity was 96.60 (SD = 4.83).

Outcomes

Table 2 presents summary data for each outcome at each data point (T1, T2, T3). There was minimal missing data for the face-to-face assessments. However, the FOCUS-34 returns from parents were very low (T1: 68/102; T2: 49/102; T3: 43/102) and so are not used for further analyses. Returns from teachers were more complete.

----------Table 2. Outcome measures and group comparisons ----------

Results of the ANCOVA analyses are presented in Table 2 and highlighted in bold where significant differences between time points and between groups were found. Group comparisons at T2 and T3 were adjusted for T1 scores. The following interprets those results with reference to each research question. Effect sizes (ES) are also presented in Table 2. Cohen's d was calculated from ANCOVA partial eta squared using an effect size converter (MRC Cognition and Brain Sciences Unit, 2009) based on formulae from Cohen (1988, pp. 281, 284, 285). ES are interpreted below with reference to Coe et al.'s criteria for use in educational interventions which provide an estimate of months progress gained from the intervention (Coe et al., 2013) (Appendix 5).

RQ 1 Which intervention is most effective?

BEST was associated with significantly greater gains than A-DLS in SS in NRDLS comprehension between T2 and T3 with a high ES (.56), and production between T1 and T3, again with high ES (.55). Gains between T2 and T3 for production SS did not reach significance (p = .059; moderate ES = .40) however when the confound of wave is included this comparison is significant. BEST was also associated with greater gains in BEST targeted sentences at T2 and T3 with very high (.78) and high (.44) effect sizes respectively. There were no other significant differences between the treatment arms (see Figures 2-4 and Table 2).

RQ 2 Do interventions differ in the degree to which benefits transfer to non-targeted language structures and/or communicative participation?

Both intervention arms made significant improvements from T1 to T2 in all outcomes: NRDLS Production SS, NRDLS Comprehension SS, BEST Assessment scores, A-DLS Adapted RST and FOCUS-34. These benefits were maintained at T3 such that significant improvements from T1 to T3 were also present for all measures.

Improvements in NRDLS SS, which are corrected for age, potentially represent catch-up growth in language with changes in average standard scores from T1 to T2 of 8 for comprehension (BEST: 5; A-DLS: 9) and 9 for production (BEST: 10; A-DLS: 7); and between T1 and T3 of 11 for comprehension (BEST: 12; A-DLS: 9) and 9 for production (BEST: 13; A-DLS: 7). The FOCUS-34 classifies a change in score of > 16 as a significant clinical change (SCC) in communicative participation. The average change between T1 and T2 was 33 (BEST: 32; DLS: 34) with 66% of children reaching the SCC threshold (BEST: 73%; A-DLS – 61%). At T3 there was substantial missing data and so T1 to T3 changes are not considered further (see Table 2).

Results on NRDLS and FOCUS-34 represent non-targeted language structures and communicative participation respectively. BEST was associated with significantly greater gains in both NRDLS

comprehension and production than for A-DLS. There were no significant differences between groups in communicative participation (Table 2 Figures 2-4).

RQ 3 Do interventions differ in the degree to which language abilities continue to improve after the intervention is complete?

BEST intervention was associated with significantly greater gains between T2 and T3 than A-DLS in NRDLS comprehension SS and production SS in an adjusted model (see below). There were no significant differences between groups on T2 to T3 gains for the measures of targeted language structures (BEST Assessment and A-DLS Adapted RST). However, the A-DLS was associated with significant progress from T2 – T3 in the A-DLS Adapted Rapid Screening Test which was not present for those receiving the BEST intervention.

Potential confounds

Finally, models examining NRDLS comprehension and production SS were rerun adjusted for the potential confounds of wave and delayed intervention. Due to the small sample size these were completed in separate models. Results are presented in Appendix 4. No substantive differences in the pattern of results were found although the difference between BEST and A-DLS for NRDLS T2-T3 gains in production SS becomes significant when wave was covaried.

Discussion

Both interventions were associated with significant improvements in all outcomes, including in Standard Scores, although in the absence of a no-treatment control such changes alone cannot be interpreted as proof of efficacy. However, comparison between two active intervention arms provide a highly stringent test of efficacy should, as in this case, one arm yield greater gains than the other.

Most of the study hypotheses were supported. As predicted, BEST was associated with greater gains in children's production with high effect sizes (ES) (d= .55); however, we did not find the predicted greater gains for comprehension for A-DLS, where the interventions appeared to bring equivalent benefits. Due to the hypothesised promotion of abstract representations allowing knowledge to be generalised we expected BEST would bring greater benefits in non-targeted structures and promote greater gains after the intervention. This was supported by greater gains in production Standard Scores and by the pattern of accelerated progress after the intervention for both production and comprehension scores. That is, BEST was associated with greater gains after the intervention was complete (T2–T3) with moderate to high ES (Comprehension d = .56; Production d = .40). Raw score gains in targeted language structures (BEST Assessment and A-DLS Adapted RST) between baseline and outcome (T1-T2) favoured BEST with a very high ES for BEST Assessment scores (d = .77) and no significant differences were found between interventions in the A-DLS Adapted RST scores. BEST is, therefore, effective in improving the production of sentence structures targeted in the intervention and in promoting generalisation to non-targeted language structures. The interventions are equally effective at improving comprehension.

The majority of children across interventions (66%) made gains in their communicative participation which reached the threshold for a clinically significant change and there were significant changes for both interventions between T1 and T2 in this outcome measure. There were no significant group differences in this outcome. These gains in both language SS and communicative participation outcomes are very encouraging particularly given the low dosage of sixteen 15-minute small-group interventions over 8 weeks.

The ES described above are interpreted with reference to Coe et al.'s criteria for use in educational interventions (Coe et al., 2013) which were developed for comparison between an intervention and TAU control. This study, comparing two active interventions, therefore provides a highly conservative estimate of ES yet we find larger effects than other targeted small-group interventions

with higher dosage (Bleses et al., 2018; West et al., 2024). These positive findings in the relatively socially disadvantaged sample in the study are particularly encouraging given that some preschool interventions can widen rather than narrow inequalities (McKean & Reilly, 2023).

The lack of a non-intervention control due to changes to the study following COVID restrictions makes it difficult to be sure that A-DLS brings benefits over and above usual EYS practice. However, previous research would suggest this is highly likely (Broomfield & Dodd, 2011). Creating change in non-targeted language structures is vital for effective and efficient intervention. Greater gains for the BEST intervention in production SS, that is language structures not targeted in the intervention, prove BEST's efficacy, supporting the findings of Trebacz et al. (2023) but in a more rigorous study methodology. Furthermore, these findings suggest it may be more efficient than other interventions although this requires further research.

Faster progress after intervention supports the hypothesis that the active ingredients in BEST, based on usage-based theory, do promote the development of abstract representations of PAS, supporting generalisation and accelerating language learning. Longer-term follow-up is needed to test how long such benefits might be present for a child; however, it demonstrates the significant promise of interventions designed with reference to usage-based theory and which precisely manipulate hypothesised cognitive and linguistic active ingredients to promote change.

Strengths and Limitations

This preregistered (ISRCTN10974028) cluster randomised controlled trial following CONSORT guidance for conduct and reporting represents a rigorous evaluation of the relative efficacy of BEST and A-DLS including randomisation, blinding, extensive treatment fidelity strategies, and minimal dropout and missing data (Campbell et al., 2012). The effects of COVID meant the study had a smaller overall sample size than originally planned. The sample in each treatment arm did not reach the target of sixty-five derived from power calculations (A-DLS = 58; BEST = 45); however, data were

maximised through high retention and data completeness. The smaller sample size could inflate the effect sizes found.

We acknowledge that the analysis performed does not take into account the non-independence in the data (child within school), and that a multilevel approach would have been preferable. This approach was not adopted given the numbers of schools and children in each school (in some cases only a single child) (Maas & Hox, 2005; McNeish, 2014). Bootstrapping was used with no substantive differences obtained.

The A-DLS Adapted RST used to assess progress in target structures did contain some structures not targeted in the A-DLS for each child and so perhaps could underestimate the progress made in this arm.

Removing the TAU arm due to ethical concerns made it harder to draw conclusions about the efficacy of whichever intervention had a smaller effect, in this case, A-DLS. Furthermore, the adaption of A-DLS makes it difficult to generalise our findings to standard DLS. However, this head-to-head approach, matching interventions with respect to dosage, delivery context, fidelity, etc., enables us to rule out general therapy effects. We can be sure that it is the specific qualities of BEST that effect change in production scores over and above small group play activities.

Conclusions

Services need access to multiple interventions with proven efficacy to choose approaches that best fit the needs of the populations they serve, align with the constraints of service provision, and bring lasting benefits to children. This study provides evidence from a rigorous RCT that it is possible to bring about moderate to high language gains for preschool children, including those from low socio-economic backgrounds, with a low dosage intervention (d = .44 BEST assessment raw scores; d = .55 NRDLS production SS). Our findings using a head-to-head comparator likely underestimate these effects, which Coe et al.'s approach suggests are equivalent to 5 and 7 months progress,

respectively. Given the intense pressures on early years and SLT services, it is essential that effective interventions are identified which are feasible for implementation within such constraints. BEST may offer such an approach for some children. In terms of profiles of need this study shows that children with production difficulties benefit from BEST. Previous study findings suggest DLS improves comprehension. It is, therefore, possible that BEST and DLS both bring benefits in this domain, but further research is required to test this assumption regarding BEST.

The accelerated progress after intervention underscores the significant potential of interventions designed with reference to usage-based theory and which precisely manipulate language exposure to leverage cognitive mechanisms to promote language learning and abstract knowledge. We recommend further research is conducted to examine whether longer-term gains persist after intervention, the wider potential of the application of usage-based theory to language interventions, and health economic evaluation to consider both the efficacy and efficiency of early interventions.

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

Axford, N., Sonthalia, S., Wrigley, Z., Goodwin, A., Ohlson, C., Bjornstad, G., Barlow, J., Schrader-McMillan, A., Coad, J., & Toft, A. (2015). *The best start at home*. Early Intervention Foundation.

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

Bloom, L., & Lahey, M. (1978). *Language development and language disorders*. Wiley.

Bloom, L., Lightbown, P., Hood, L., Bowerman, M., Maratsos, M., & Maratsos, M. P. (1975). Structure and variation in child language. *Monographs of the Society for Research in Child Development*, 1-97.

Broomfield, J., & Dodd, B. (2011). Is speech and language therapy effective for children with primary speech and language impairment? Report of a randomised control trial. *International Journal of Language & Communication Disorders*, *46*(6), 628-640.

Brown, R. (1976). *A first language: the early stages*. Penguin Books.

Brown, R., Cazden, C. B., & Bellugi-Klima, U. (1981). The child's grammar from I to III. In C. A. Ferguson (Ed.), *Minnesota Symposia on Child Psychology* (1 ed., Vol. 2). Psychology Press.

Campbell, M. K., Piaggio, G., Elbourne, D. R., & Altman, D. G. (2012). Consort 2010 statement: extension to cluster randomised trials. *BMJ : British Medical Journal*, *345*, e5661. https://doi.org/10.1136/bmj.e5661

Coe, R., Kime, S., Neville, C., & Coleman, R. (2013). *The DIY evaluation guide London: Education Endowment Foundation.* Education Endowment Foundation.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Routledge Academic.

Eadie, P., Levickis, P., Murray, L., Page, J., Elek, C., & Church, A. (2021). Early Childhood Educators' Wellbeing During the COVID-19 Pandemic. *Early Child Educ J*, *49*(5), 903-913. https://doi.org/10.1007/s10643-021-01203-3

Early Years Alliance. (2021). *Breaking point: The impact of recruitment and retention challenges on the early years sector in England*. E. Y. Alliance. https://www.eyalliance.org.uk/sites/default/files/breaking_point_early_years_alliance_2_december_2021.pdf

Education and Skills Funding Agency. (2022). *Pupil premium: conditions of grant 2021 to 2022 for academies and free schools*. https://www.gov.uk/government/publications/pupil-premium-allocations-and-conditions-of-grant-2021-to-2022/pupil-premium-conditions-of-grant-2021-to-2022-for-academies-and-free-schools

Edwards, S., Letts, C. A., & Sinka, I. (2011). *New Reynell Developmental Language Scales (NRDLS)*. GL Assessment.

Frizelle, P., & McKean, C. (2022). Using Theory to Drive Intervention Efficacy: The Role of Dose Form in Interventions for Children with DLD [Article]. *Children*, *9*(6), Article 859. https://doi.org/10.3390/children9060859

Frizelle, P., Mullane, E., O'Shea, A., Ceroni, A., Dahly, D., Horgan, A., Levickis, P., & Mckean, C. (2021). Happy Talk: A pilot effectiveness study of a targeted-selective speech–language and communication intervention for children from areas of social disadvantage. *International Journal of Language & Communication Disorders*, *56*(5), 954-974. https://doi.org/https://doi.org/10.1111/1460-6984.12648

Frizelle, P., Tolonen, A. K., Tulip, J., Murphy, C. A., Saldana, D., & McKean, C. (2021a). The impact of intervention dose form on oral language outcomes for children with developmental

language disorder [Review]. *Journal of Speech, Language, and Hearing Research*, *64*(8), 3253-3288. https://doi.org/10.1044/2021_JSLHR-20-00734

Frizelle, P., Tolonen, A. K., Tulip, J., Murphy, C. A., Saldana, D., & McKean, C. (2021b). The influence of quantitative intervention dosage on oral language outcomes for children with developmental language disorder: A systematic review and narrative synthesis [Review]. *Language, Speech, and Hearing Services in Schools*, *52*(2), 738-754. https://doi.org/10.1044/2020_LSHSS-20-00058

Greenwood, C. R., Schnitz, A. G., Carta, J. J., Wallisch, A., & Irvin, D. W. (2020). A systematic review of language intervention research with low-income families: A word gap prevention perspective. *Early Childhood Research Quarterly*, *50*, 230-245. https://doi.org/10.1016/j.ecresq.2019.04.001

Hagen, A. M., Melby-Lervag, M., & Lervag, A. (2017). Improving language comprehension in preschool children with language difficulties: a cluster randomised controlled trial. *Journal of Child Psychology and Psychiatry*, *58*(10), 1132-1140.

Hulme, C., Nash, H. M., Gooch, D., Lervåg, A., & Snowling, M. J. (2015). The Foundations of Literacy Development in Children at Familial Risk of Dyslexia [Article]. *Psychological Science*, *26*(12), 1877-1886. https://doi.org/10.1177/0956797615603702

Knowles, W., & Masidlover, M. (1982). *The Derbyshire Language Scheme*. Derbyshire County Council.

Langacker, R. W. (2000). A dynamic usage-based model. In M. Barlow & S. Kemmer (Eds.), *Usage-based models of language* (pp. 1-63). CSLI Publications.

Law, J., & Charlton, J. (2022). Interventions to Promote Language Development in Typical and Atypical Populations. In L. J, S. Reilly, & C. McKean (Eds.), *Language Development: Individual Differences in a Social Context* (pp. 470-494). Cambridge University Press. https://doi.org/10.1017/9781108643719.024

Law, J., Charlton, J., Dockrell, J., Gascoigne, M., McKean, C., & Theakston, A. (2017). *Early Language Development: Needs, provision, and intervention for preschool children from socio-economically disadvantage backgrounds.* Institute of Education.

Law, J., Rush, R., Schoon, I., & Parsons, S. (2009). Modeling developmental language difficulties from school entry into adulthood: Literacy, mental health, and employment outcomes. *Journal of Speech Language and Hearing Research*, *52*(6), 1401-1416. http://www.scopus.com/inward/record.url?eid=2-s2.0-71649093381&partnerID=40&md5=451f9bd3152f97ef794a88f44e9d12f1

Le, H. N. D., Mensah, F., Eadie, P., McKean, C., Sciberras, E., Bavin, E. L., Reilly, S., & Gold, L. (2021). Health-related quality of life of children with low language from early childhood to adolescence: results from an Australian longitudinal population-based study. *Journal of Child Psychology and Psychiatry*, *62*(3), 349-356. https://doi.org/https://doi.org/10.1111/jcpp.13277

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. . *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *1*(3), 86-92. https://doi.org/doi:10.1027/1614-2241.1.3.86.

Mansournia, M. A., Nazemipour, M., Naimi, A. I., Collins, G. S., & Campbell, M. J. (2020). Reflection on modern methods: demystifying robust standard errors for epidemiologists. *International Journal of Epidemiology*, *50*(1), 346-351. https://doi.org/10.1093/ije/dyaa260

McKean, C., Benson, K., Jack, C., Letts, C. A., Pert, S., Preston, E., Trebacz, A., Stringer, H., & Wareham, H. (2020). *ISRCTN10974028: Language intervention in the Early Years - comparing the effectiveness of language intervention approaches for pre-school children with language difficulties.* https://doi.org/10.1186/ISRCTN10974028.

McKean, C., Pert, S., & Stow, C. (2013). *Building Early Sentences Therapy: A home language, early intervention programme for young children with severe language difficulties*. Newcastle University.

McKean, C., & Reilly, S. (2023). Creating the conditions for robust early language development for all: Part two: Evidence informed public health framework for child language in the early years. *International Journal of Language & Communication Disorders*, *n/a*(n/a). https://doi.org/https://doi.org/10.1111/1460-6984.12927

McNeish, D. M. (2014). Modeling sparsely clustered data: Design-based, model-based, and single-level methods. *Psychological Methods*, *19*(4), 552-563. https://doi.org/10.1037/met0000024

Ministry of Housing Communities and Local Government. (2019). *English indices of deprivation 2019*. https://imd-by-postcode.opendatacommunities.org/imd/2019

MRC Cognition and Brain Sciences Unit. (2009). *Computing effect sizes*. Retrieved 20/10/2023 from https://imaging.mrc-cbu.cam.ac.uk/statswiki/FAQ/Escomp?highlight=%28converter%29

Reilly, S., & McKean, C. (2023). Creating the conditions for robust early language development for all: Part 1: Evidence informed child language surveillance in the early years. *International Journal of Langauge and Communication Disorders*. https://doi.org/https://doi.org/10.1111/1460-6984.12929

Reilly, S., Tomblin, B., Law, J., McKean, C., Mensah, F., Morgan, A., Goldfeld, S., Nicholson, J., & Wake, M. (2014). SLI: a convenient label for whom? *International Journal of Language & Communication Disorders*, *49*(4), 416-451.

Roulstone, S., Wren, Y., Bakopoulou, I., Goodlad, S., & Lindsay, G. (2012). *Exploring interventions for children and young people with speech, language and communication needs: A study of practice* Department for Education.

Rowe, J. (1981). The Paget-Gorman Sign System. *British Journal of Special Education*, *8*(4), 25-27.

Schoon, I., Parsons, S., Rush, R., & Law, J. (2009). Children's language ability and psychosocial development: A 29-year follow-up study. *Pediatrics*, *126*(1). http://www.scopus.com/inward/record.url?eid=2-s2.0-77954378607&partnerID=40&md5=b90e3ac51c591a239b6a93c3dba52b82

Schoon, I., Parsons, S., Rush, R., & Law, J. (2010). Childhood language skills and adult literacy: A 29-year follow-up study. *Pediatrics*, *125*(3), e459-e466. https://doi.org/10.1542/peds.2008-2111

Snowling, M. J., West, G., Fricke, S., Bowyer-Crane, C., Dilnot, J., Cripps, D., Nash, M., & Hulme, C. (2022). Delivering language intervention at scale: promises and pitfalls. *Journal of Research in Reading*, *45*(3), 342-366. https://doi.org/https://doi.org/10.1111/1467-9817.12391

Sparrow, S., Cicchetti, D., & Saulnier, C. (2016). Vineland Adaptive Behavior Scales–Third Edition: Manual. *Bloomington, MN: Pearson*.

Tomasello, M. (2000). First steps toward a usage-based theory of language acquisition. *Cognitive Linguistics*, *11*(1-2), 61-82. <Go to ISI>://000167900400005

Tomasello, M. (2003). *Constructing a language: A usage based theory of language acquisition*. Harvard University Press.

Tomblin, J. B. (2014). Educational and psychosocial outcomes of language impairment in kindergarten. In *Understanding Individual Differences in Language Development Across the School Years* (pp. 166-203). https://doi.org/10.4324/9781315796987

Törnqvist, M. C., Thulin, S., Segnestam, Y., & Horowitz, L. (2009). Adult people with Language Impairment and their life situation. *Communication Disorders Quarterly*, *30*(4), 237-254. https://doi.org/10.1177/1525740108326034

Tracey, L., Bowyer-Crane, C., Bonetti, S., Nielsen, D., D'Apice, K., & Compton, S. (2022). The Impact of the COVID-19 Pandemic on Children's Socio-Emotional Wellbeing and Attainment during the Reception Year. Research Report. *Education Endowment Foundation*.

Trebacz, A., McKean, C., Pert, S., & Stringer, H. (2023). Piloting Building Early Sentences Therapy for pre-school children with low language abilities: an examination of efficacy and the role of sign as an active ingredient. *International Journal of Language and Communication Disorders 00*, 1-24.

Walker, M. (1987). *The Makaton Vocabulary- Uses and Effectiveness* International Afasic Symposium of Specific Language Disorders in Children, Reading.

West, G., Lervåg, A., Birchenough, J. M. H., Korell, C., Rios Diaz, M., Duta, M., Cripps, D., Gardner, R., Fairhurst, C., & Hulme, C. (2024). Oral language enrichment in preschool improves children's language skills: a cluster randomised controlled trial [Article]. *Journal of Child Psychology and Psychiatry and Allied Disciplines*. https://doi.org/10.1111/jcpp.13947

Table 1. Early Years Setting and Child Participant Characteristics

	All settings		BEST		DLS		*p*
	Wave 1	Wave 2	Wave 1	Wave 2	Wave 1	Wave 2	
Early Years Setting							
N	10	10	5	5[†]	5	5	
SES - Ever 6 [‡] M (SD) range	37.50 (20.31) 11.10 – 63.50	38.57 (20.16) 11.00 – 64.00	41.58 (21.92) 11.10 – 63.50	34.18 (20.95) 11.00 – 53.70	33.42 (20.16) 15.50 – 60.90	42.08 (21.20) 18.50 – 64.00	0.95
Children							
N	50	52	21	23	29	29	0.23
No. per setting M (SD) range	5 (3.16) 1-10	5.2 (1.87) 1 - 7	4.2 (2.17) 2 - 7	4.6 (2.51) 1 - 7	5.8 (4.02) 1 - 10	5.8 (0.84) 5-7	
Gender %M/F	56/44	60/40	43/57	61/39	66/34	59/41	0.32
Age M (SD) range	4;00 (0;03) 3;06 – 4;05	3;11 (0;04) 3;05 – 4;05	3;10 (0;02) 3;07 – 4;05	3;10 (0;03) 3;05 – 4;04	4;00 (0;03) 3;06 – 4;05	3;11 (0;04) 3;05 – 4;05	0.30
NRDLS Comprehension SS M (SD) range	79.88 (7.79) 69-95	79.90 (9.59) 69-116	83.38 (7.53) 69-95	82.96 (11.65) 69-116	77.34 (7.06) 69-92	77.48 (6.87) 69-92	0.001
NRDLS Production SS M (SD) range	73.62 (5.17) 69-86	76.79 (9.02) 69-107	74.24 (5.00) 69-85	76.61 (9.04) 69-101	73.17 (5.33) 69-86	76.93 (9.16) 69-107	0.78
Vineland SS [§] M (SD) range	82.68 (6.42) 68-97	83.28 (9.85) 69-107	83.72 (4.94) 78 - 97	84.43 (10.37) 69 - 107	82.03 (7.19) 68 – 95	81.67 (9.19) 71 - 103	0.22
Group sizes M (SD); range	2.4 (1.1) 1-5	2.3 (1.1) 1-5	3.0 (0.6) 2-4	2.6 (1.3) 1-5	2.1 (1.2) 1-5	2.1 (1.1) 1-4	0.05
SES IDACI Median, IQR range	1 (2)	1.5 (3)	2 (4)	2 (3)	1 (2)	1 (3)	0.12

Notes *p* values represent comparisons between intervention arms for each demographic characteristic using t-tests, independent samples median test or Chi-squared as appropriate. Key: SES = socioeconomic status; Ever 6 = used to calculate pupil premium is the number of children in a setting who had a recorded period of free school meals in the previous 6 years (Education and Skills Funding Agency, 2022); NRDL = New Reynell Developmental Language Scales; SS = Standard scores; Vineland = a measure of adapted behaviour in three domains: communication, daily living skills and socialisation and combine to provide an Adaptive Behaviour Composite. (Sparrow et al., 2016); IDACI: a composite index of deprivation for postcode in England which are ranked (Ministry of Housing Communities and Local Government, 2019)[†] Ever6data available for 4 settings at wave 1[‡] Ever6 data is not available for early years settings which are not part of a school, 1 setting was a standalone nursery. This data is provided for 10 settings in wave 1 and 9 for wave 2 [§] n= 83

Table 2. Number of completed assessments (N) Mean (M) and Standard Deviation (SD) scores for the BEST and A-DLS interventions at preintervention (T1) post-intervention (T2) and follow-up (T3), ANCOVA results examining within and between-group differences with T1 scores as covariate.

		Full sample			BEST Intervention			A-DLS Intervention			d/η^2	BEST - A-DLS Model		Within arm BEST		Within arm DLS	
		N	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*		*F (df)*	*P*	*t (df)*	*p*	*t (df)*	*P*
NRDLS Comprehension SS																	
	T1	102	78.89	8.71	44	83.16	9.79	58	77.41	6.91							
	T2	100	87.22	12.89	44	88.45	13.23	56	86.25	12.65							
	T3	102	90.23	15.3	44	94.89	14.23	58	86.69	15.24							
	T1 – T2										.06/.001	0.103(1)	.748	-2.52(43)	.016	-5.61(55)	<.001
	T2 – T3										.56/.07	7.593(1)	**.007**	-4.17(43)	<.001	-.56(55)	.577
	T1 – T3										.26/.02	1.631(1)	.205	-5.81(43)	<.001	-5.30(57)	<.001
NRDLS Production SS																	
	T1	102	75.24	7.52	44	75.48	7.41	58	75.05	7.66							
	T2	95	84.08	13.09	44	86.16	12.98	51	82.29	13.05							
	T3	102	84.98	12.94	44	88.7	13.96	58	82.16	11.44							
	T1 – T2										.30/.02	2.16(1)	.145	-6.75(43)	<.001	-5.24(55)	<.001
	T2 – T3										.40/.04	3.64(1)	.059	-1.47(43)	.149	-0.12(50)	.903
	T1 – T3										.55/.07	7.56(1)	**.007**	-71(43)	<.001	-5.41(57)	<.001
BEST Assessment																	
	T1	98	26.12	16.82	41	29.21	18.70	57	23.89	15.11							
	T2	100	57.41	22.49	44	67.93	23.47	56	49.14	17.93							
	T3	102	56.82	21.91	44	64.29	21.51	58	51.16	20.63							
	T1 – T2										.77/.13	13.864(1)	**<.001**	-10.24(40)	<.001	-11.33(54)	<.001
	T2 – T3										.00/.00	0.04(1)	.841	1.26(43)	.215	-0.87(55)	.390
	T1 – T3										.44/.05	4.6(1)	**.035**	-9.67(40)	<.001	-10.58(56)	<.001
A-DLS Adapted RST																	
	T1	102	47.2	8.08	44	49.61	7.07	58	45.36	8.38							
	T2	100	51.45	7.22	44	53.2	4.91	56	50.07	8.41							
	T3	102	52.99	5.98	44	54.27	5.17	58	52.02	6.4							
	T1 – T2										.21/.01	1.09(1)	.299	-4.05(43)	<.001	-3.89(55)	<.001
	T2 – T3										.13/.004	0.38(1)	.541	-1.79(43)	.08	-2.17(55)	.034
	T1 – T3										.14/.01	0.48(1)	.489	-4.68(43)	<.001	-6.55(55)	<.001
FOCUS-34 score																	
	T1	98	132.47	44.39	44	132.89	38.17	54	132.13	49.24							
	T2	93	166.22	39.37	40	165.73	35.64	53	166.58	42.31							
	T3	77	162.57	41.21	23	154.61	41.6	54	165.96	40.96							
	T1 – T2										.17/.01	0.58(1)	.447	-6.51(39)	<.001	-6.98(50)	<.001
FOCUS-34 SCC			% SCC			% SCC		*N*	% SCC			Chi^2	*P*				
	T1 – T2	93	66			73			61			1.50	.22				

Key: NRDLS: New Reynell Developmental Language Scales; SS: Standard Score; BEST: Building Early Sentences Therapy; A-DLS: Adapted Derbyshire Language Scheme; RST: Rapid Screening Test; FOCUS-34: Functional Outcomes in Children Under 6; SCC Significant Clinical Change

Figure 1. CONSORT diagram showing participant flow through the study

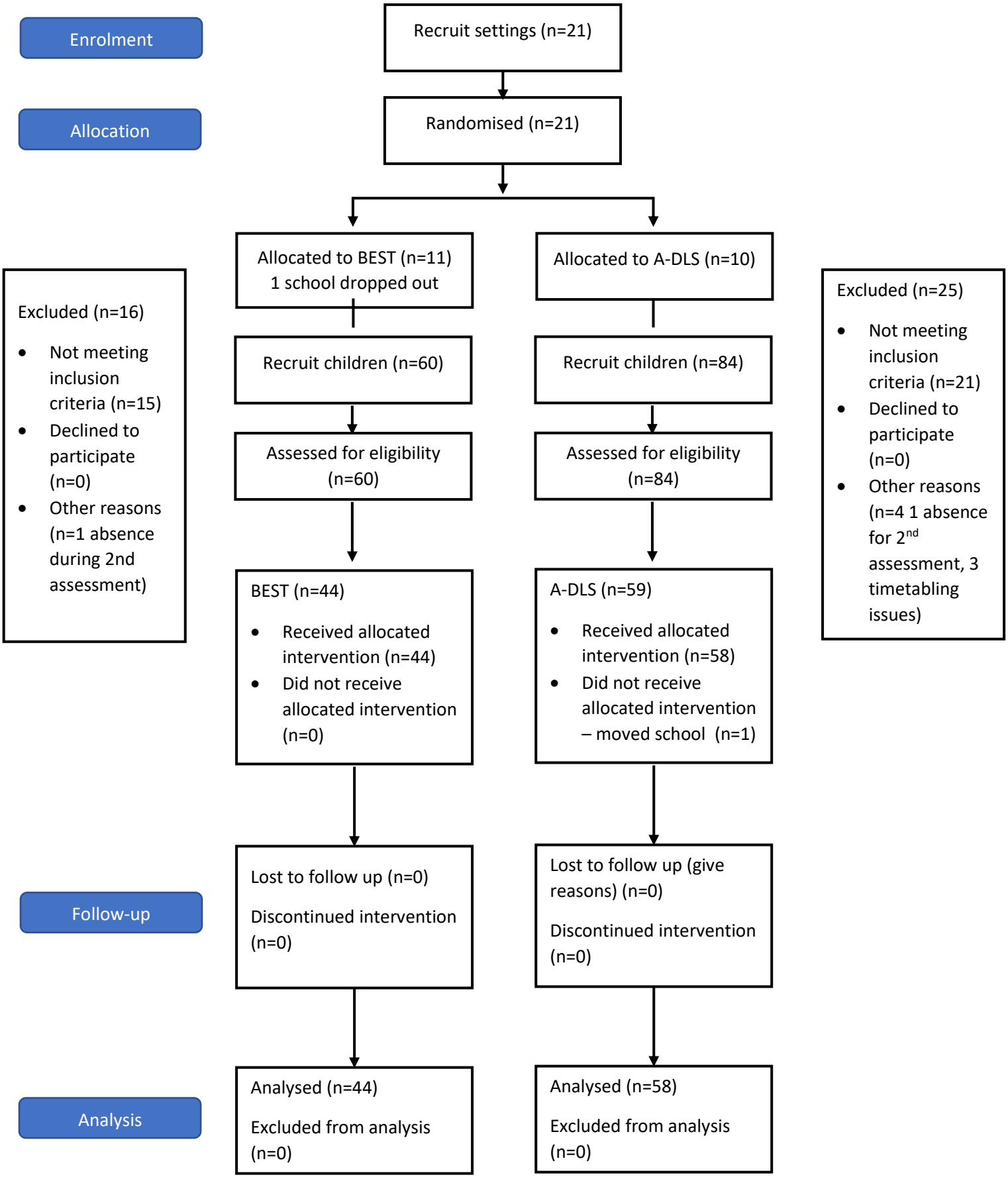
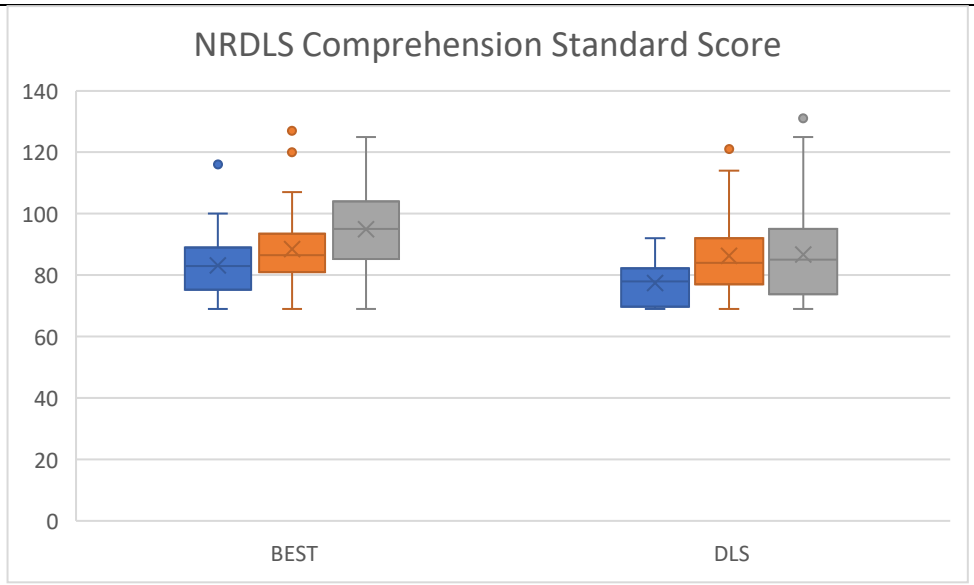


Figure 2 Box and whisker plots of NRDLS Comprehension and Production Standard Score at T1, T2 and T3 for each intervention arm



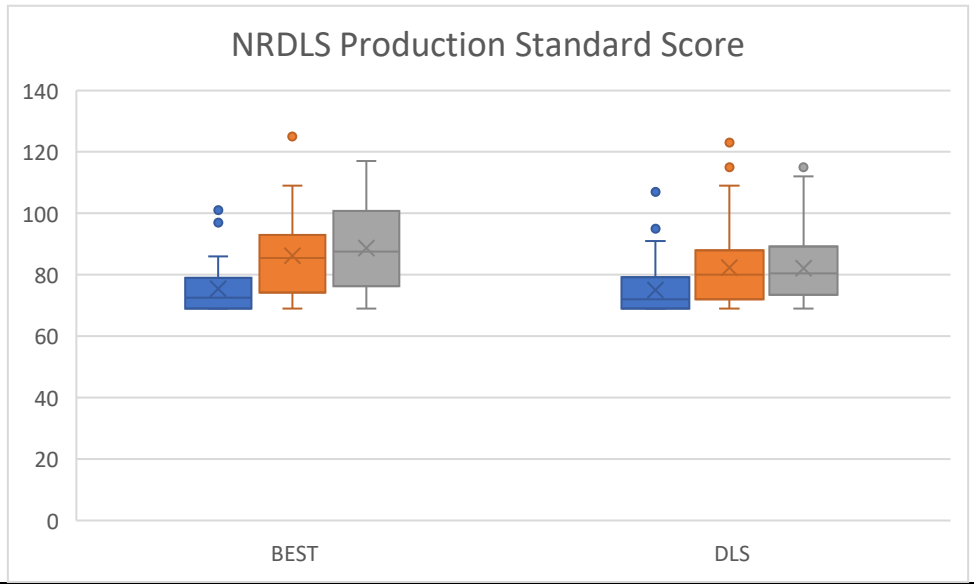


Figure 3 Box and whisker plots of BEST Assessment and A-DLS Adapted RST Scores at T1, T2 and T3 for each intervention arm

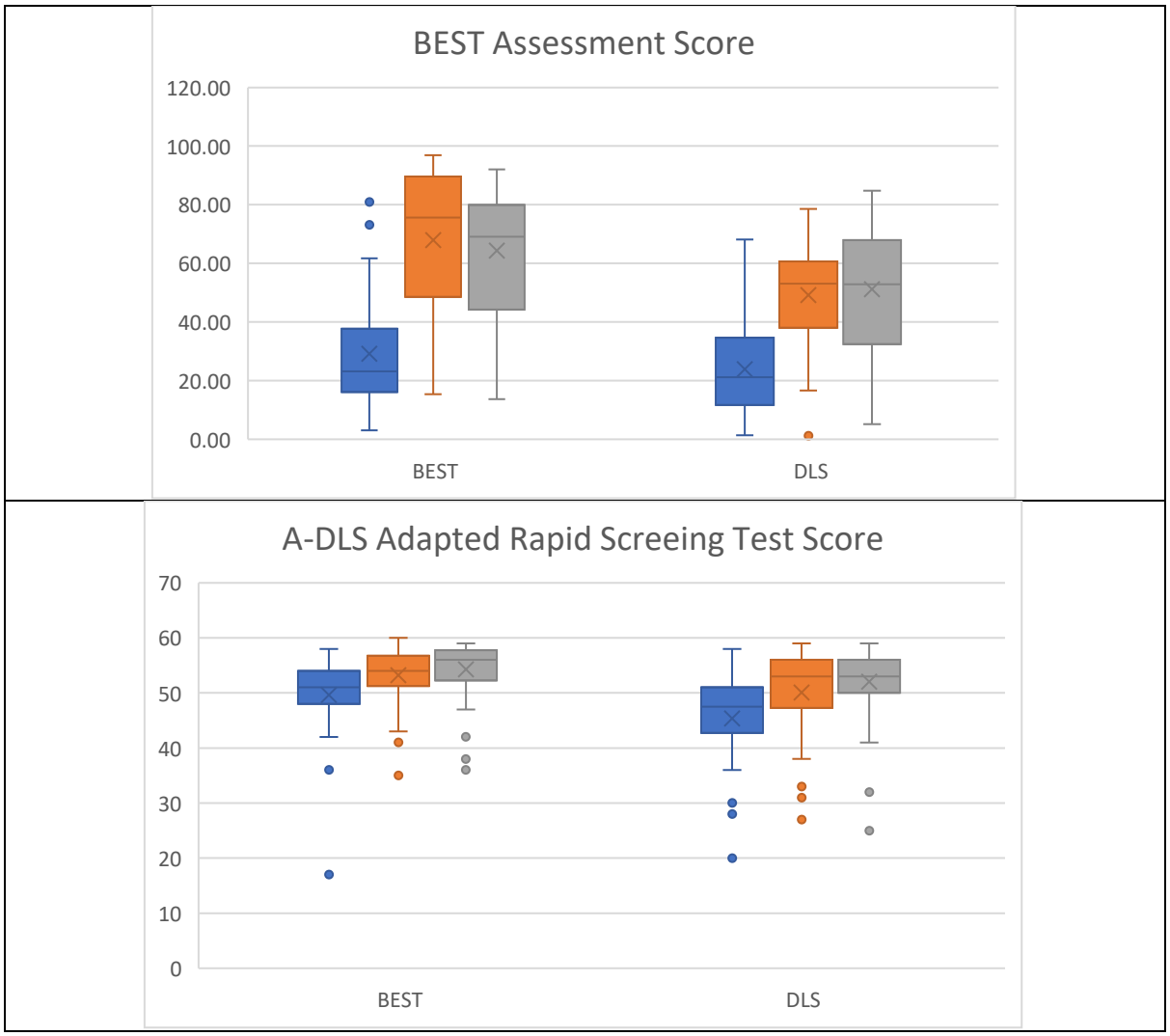
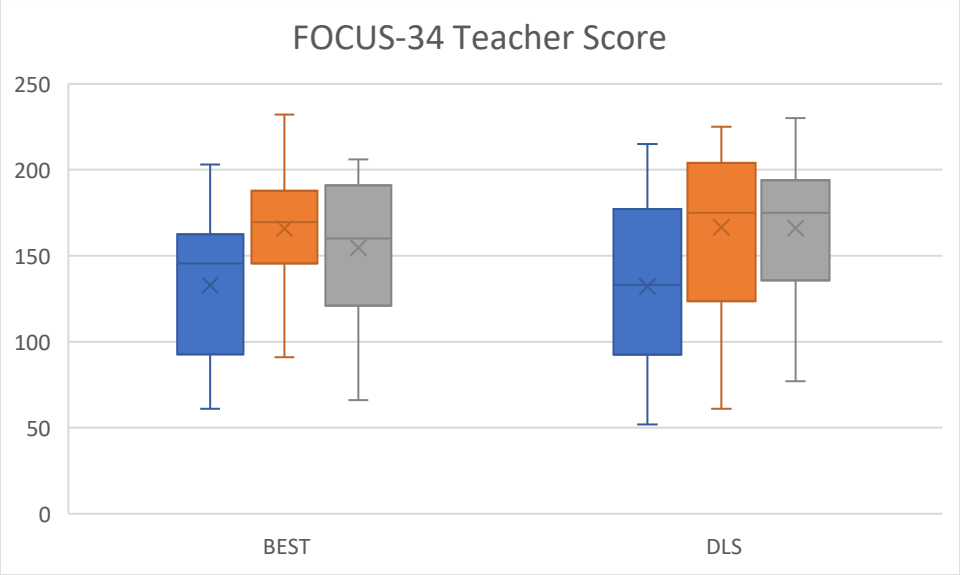


Figure 4 Box and whisker plot of FOCUS-34 at T1, T2 and T3 for each intervention arm



Appendix 1 - Stages of multi-word utterance development, relevant cognitive mechanisms, and their use as active ingredients in BEST (McKean et al., 2013; Tomasello, 2003). Reproduced from Trebacz, A., McKean, C., Stringer, H. & Pert, S. (2024) Piloting building early sentences therapy for pre-school children with low language abilities: An examination of efficacy and the role of sign as an active ingredient. *International Journal of Language & Communication Disorders*, 59, 1128–1151. https://doi.org/10.1111/1460-6984.12980

Stage	Relevant cognitive mechanisms		
	Name	Description	Relevant BEST 'active ingredient'
1. Frozen Phases Rote learned, and therefore inflexible, utterances paired with a pragmatic function and a communicative context/cultural routine. The child cannot combine the elements of the structure productively with other words. e.g. "*eat it*" bound to a meal time social routine or "*shoes off"* bound to an undressing routine	*Intention reading*	To create the form-function mapping required for the development of frozen phrases, the child must 'read' the communicative intentions of the person from whom they are learning the phrase. The child's ability to read the intentions of others within the scaffolding of joint attentional frames (Tomasello, 2003 p.21).	A structured and repetitive 'joint action routine' is established, creating a joint attentional frame between the child and the adult, which 'scaffolds' the child's ability to infer the communicative intention of the utterances they hear (i.e. describing an event within a play activity). Hence the child quickly becomes able to infer the communicative intentions of the adult at the level of the attentional frame (which objects and actions are we both attending to and what is the global purpose of this joint attention); and so is supported to make such inferences at the level of the individual communicative acts within that frame (which objects and actions is the adult referring to with a specific utterance) (Tomasello, 2003).
	Cultural learning	A process by which young children learn through imitation (and later through instructed and collaborative learning) of others in their social group (Tomasello, 2003 p.290). The child must not only mirror the communicative behaviour of the adult, but also understand that the roles within the triadic attentional frame (e.g. adult - child - object/action) reverse when they imitate the adult. Through this process, the child	Role reversal is used within the 'joint action routine' to promote cultural learning and hence the creation of symbolic linguistic representations (Tomasello, 2003).

Stage	Relevant cognitive mechanisms		
	Name	Description	Relevant BEST 'active ingredient'
		comprehends that when they are the speaker, imitating the communication of the adult, the communicative intention of the adult that was directed to the child instead becomes directed to the adult (Tomasello, 2003 p.26).	
2. *Lexically specific constructions* Partially productive/flexible utterances with a 'slot and frame' construction where only one element can vary (e.g. "*X fall down*" or "*I'm ACTIONing it*"). 'Frame' categories might be: "X fall down" – the category of events in which animate or inanimate objects	*Schemat-isation*	A general cognitive strategy that facilitates the identification of rules and patterns or *schemas*, or within the child's environment, supporting them to rely on mental abstractions (Piaget, 1952). In the case of communication, multiple exposures to the same utterances where one component is varied across exposures (e.g. X fall down, where X is *the girl* and then *the boy* and then *the teddy)* enable children to create rules or *schemas* which represent the aspects of the construction that remain the same across iterations (the 'frame'), and which components vary across iterations (the 'slot', in this case X) (Gomez, 2002; Tomasello, 2003 p.122).	The cognitive processes of *schematisation* and *categorisation* both depend on the quantity and distribution of types and tokens within the input heard by the child. BEST provides multiple presentations of highly similar exemplar sentences in which one element is varied systematically (Gomez, 2002; Tomasello, 2003). (e.g. The baby is laughing; The woman is laughing; The girl is laughing; The teddy is laughing).

Stage	Relevant cognitive mechanisms		
	Name	Description	Relevant BEST 'active ingredient'
unintentionally drop to a lower place "I'm ACTIONing it" – the category of events in which the child is performing an action on an object. 'Slot' categories might be: X fall down – where X is the category of animate or inanimate objects which can fall I'm ACTIONing it – where ACTION is the category of the things I can do to objects Eat X – where X is the category of objects which can be eaten.	*Categori-sation*	In order to effectively use the 'slot and frame' constructions emerging from the process of schematisation, children must also form mental *categories* of which items can be put into each 'slot'. At the stage of *lexically specific constructions*, the child's categories are still functional and relatively concrete (e.g. in the construction "*X fall down*, X might consist of 'animate objects which involuntarily move from a high place to a lower place', and the category ACTION in "*He's ACTIONing it*" would consist of 'actions 'he' can perform' (Tomasello, 2003 p.124).	

Stage	Relevant cognitive mechanisms		
	Name	Description	Relevant BEST 'active ingredient'
3. Abstract constructions Flexible, abstract representations allowing children to use any relevant lexical items in the appropriate role in the sentence and so use the sentence structures productively. Analogy across functional relationships supports the creation of semantic categories (e.g. AGENT, PATIENT) and constructions (e.g. AGENT + ACTION + PATIENT) Analogy across construction form (but not function) (e.g. ***<u>The</u>*** *girl like**<u>s</u>** cake; ***<u>The</u>*** *rabbit eat**<u>s</u>** lettuce)* results in the construction of syntactic	*Analogy*	Children identify patterns and commonalities between phenomena, in the case of linguistic abstract constructions, the shared *functional* relationships between items. Such categories are analogous because the functional relationships are the same across constructions, e.g. *'The **A** is **B**ing the **C**'* is analogous to 'The ***D** is **E**ing the **F**'* (Tomasello, 2003 p163). In the above example, **A** and **D** are doing the action, **B** and **E** are actions, and **C** and **F** are the recipients of the actions. "When an analogy is made the objects involved are effaced; the only identity they retain is their relational structure" (Tomasello, 2003 p.164).	Repeated exposure to sentence construction pairs that have the same predicate argument structure but contrasting verbs (e.g. The teddy is eating the apple; The man is washing the apple) provides children with multiple opportunities to identify the similarities in functional relationships and abstract semantic *categories* (e.g. AGENT, ACTION, PATIENT) and semantic *constructions* (e.g. AGENT + ACTION + PATIENT) (Tomasello, 2003). For each sentence construction pairing the items in each argument structure role are non-overlapping sets, providing a level of consistency thought to facilitate analogy (McKean et al., 2013) (e.g. AGENTS are never PATIENTS and vice versa). The use of toys to act out the target sentences support the identification of predicate argument structure roles (e.g. making distinctions between agent and patient more tangible). Input rotates through the different constructions targeted by BEST. This results in distributed exposure to a range of constructions across which the child can find analogies (Ambridge et al., 2006).

Stage	Relevant cognitive mechanisms		
	Name	Description	Relevant BEST 'active ingredient'
categories (e.g. VERB, OBJECT) and constructions (e.g. SUBJECT + VERB + OBJECT).			For each sentence construction pairing the morphological frame remains constant (e.g. **The** boy **is** jumping; **The** woman **is** sitt**ing; The** X **is** Y**ing the** Z) providing an additional structural cue regarding the similarity between constructions (Tomasello, 2003). The use of a signing system which marks both lexical items and grammatical morphology. The marking of these items drives pattern finding and thence analogy, supporting children with language difficulties to create abstract representations of predicate argument structure that might otherwise be difficult due to phonological and morphological processing difficulties. Sign also supports semantic mapping and reduces processing load, rendering cues in the input more accessible (Chiat, 2001; Leonard, 2007; Tomasello, 2003; Rowe, 1981; Walker & Armfield, 1981).
All stages	*Mapping*	Establishing a representation in memory of a new meaning-construction pairing which is essential for learning words and early multi-word constructions and their corresponding meanings (Hirsh-Pasek et al., 2000).	Many repetitions of the same and similar constructions are provided alongside visual referents (toys and signs) to facilitate mapping which often requires more exposures for children with language difficulties than their typically developing peers. Other verbal input is avoided (Riches et al., 2005).

Stage	Relevant cognitive mechanisms		
	Name	Description	Relevant BEST 'active ingredient'
All stages	*Retention*	The formation of robust representations of newly learned constructions in long-term memory for future retrieval (Leonard et al., 2020).	Exposure to constructions is distributed over multiple sessions to leverage spacing effects thought to facilitate long term retention of learning (Riches et al., 2005). Multiple opportunities to use the target construction expressively, facilitating long term retention (Frizelle & McKean, 2022).

Appendix 2 CONSORT CHECKLIST

Tables

Table 1 | CONSORT 2010 checklist of information to include when reporting a cluster randomised trial

Section/topic and item No	Standard checklist item	Extension for cluster designs	Page No*
Title and abstract			
1a	Identification as a randomised trial in the title	Identification as a cluster randomised trial in the title	2
1b	Structured summary of trial design, methods, results, and conclusions See table 2 (for specific guidance see CONSORT for abstracts)[1112]		2
Introduction			
Background and objectives:			
2a	Scientific background and explanation of rationale	Rationale for using a cluster design	5
2b	Specific objectives or hypotheses	Whether objectives pertain to the cluster level, the individual participant level, or both	4
Methods			
Trial design:			
3a	Description of trial design (such as parallel, factorial) including allocation ratio	Definition of cluster and description of how the design features apply to the clusters	5
3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons		6
Participants:			
4a	Eligibility criteria for participants	Eligibility criteria for clusters	5, 6
4b	Settings and locations where the data were collected		5-6
Interventions:			
5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	Whether interventions pertain to the cluster level, the individual participant level, or both	5. 8-9
Outcomes:			
6a	Completely defined prespecified primary and secondary outcome measures, including how and when they were assessed	Whether outcome measures pertain to the cluster level, the individual participant level, or both	7-8
6b	Any changes to trial outcomes after the trial commenced, with reasons		14
Sample size:			
7a	How sample size was determined	Method of calculation, number of clusters(s) (and whether equal or unequal cluster sizes are assumed), cluster size, a coefficient of intracluster correlation (ICC or *k*), and an indication of its uncertainty	5
7b	When applicable, explanation of any interim analyses and stopping guidelines		N/A
Randomisation			
Sequence generation:			
8a	Method used to generate the random allocation sequence		6
8b	Type of randomisation; details of any restriction (such as blocking and block size)	Details of stratification or matching if used	6
Allocation concealment mechanism:			
9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	Specification that allocation was based on clusters rather than individuals and whether allocation concealment (if any) was at the cluster level, the individual participant level, or both	6
Implementation:			
10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	Replaced by 10a, 10b, and 10c	6

10a		Who generated the random allocation sequence, who enrolled clusters, and who assigned clusters to interventions	6

Table 1 (continued)

Section/topic and item No	Standard checklist item	Extension for cluster designs	Page No*
10b		Mechanism by which individual participants were included in clusters for the purposes of the trial (such as complete enumeration, random sampling)	6
10c		From whom consent was sought (representatives of the cluster, or individual cluster members, or both) and whether consent was sought before or after randomisation	6
Blinding:			
11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how		7
11b	If relevant, description of the similarity of interventions		9.
Statistical methods:			
12a	Statistical methods used to compare groups for primary and secondary outcomes	How clustering was taken into account	11, 17
12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses		7
Results			
Participant flow (a diagram is strongly recommended):			
13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	For each group, the numbers of clusters that were randomly assigned, received intended treatment, and were analysed for the primary outcome	Fig 1
13b	For each group, losses and exclusions after randomisation, together with reasons	For each group, losses and exclusions for both clusters and individual cluster members	Fig 1
Recruitment:			
14a	Dates defining the periods of recruitment and follow-up		7
14b	Why the trial ended or was stopped		N/A
Baseline data:			
15	A table showing baseline demographic and clinical characteristics for each group	Baseline characteristics for the individual and cluster levels as applicable for each group	Table 1
Numbers analysed:			
16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	For each group, number of clusters included in each analysis	Table 2
Outcomes and estimation:			
17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	Results at the individual or cluster level as applicable and a coefficient of intracluster correlation (ICC or *k*) for each primary outcome	Table 2, S3
17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended		N/A
Ancillary analyses:			
18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing prespecified from exploratory		S3
Harms:			
19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms106)		
Discussion			
Limitations:			

20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses		16-17
Generalisability:			
21	Generalisability (external validity, applicability) of the trial findings	Generalisability to clusters and/or individual participants (as relevant)	-

Table 1 (continued)

Section/topic and item No	Standard checklist item	Extension for cluster designs	Page No*
Interpretation:			
22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence		15-17
Other information			
Registration:			
23	Registration number and name of trial registry		1, 5, 16
Protocol:			
24	Where the full trial protocol can be accessed, if available		N/A
Funding:			
25	Sources of funding and other support (such as supply of drugs), role of funders		Acknowledgements

*Page numbers optional depending on journal requirements.

Appendix 3

Verbs and Predicate argument structures targeted by the BEST program and the use of Contrast and Variation in those structures

No. of arguments	Set	Argument Structure	Input	Output
1	A	Agent^{2} + Action^{1}	laughing	sitting
1	B	Agent^{2} + Action^{1}	jumping	walking
2	C	Agent^{1} + Action^{1} + Patient^{2}	eating	washing
2	D	Agent^{1} + Action^{1} + Patient^{2}	riding	smelling
2	E	Agent^{1} + Action^{1} + Patient$^{2(A)}$	kissing	hugging
2	F	Agent^{1} + Action^{1} + Patient^{2}	kicking	brushing
3	G	Agent^{1} + Action + Patient^{2} + Locative^{1}	putting	putting
3	H	Agent^{1} + Action + Patient^{1} + Locative^{2}	pouring	pouring
3	I	Agent^{1} + Action^{1} + Patient$^{2(A)}$ + Locative$^{2(B)}$	putting	pouring
3	J	Agent^{1} + Action + Patient^{2} + Benefactive^{1}	giving	giving
3	K	Agent^{1} + Action + Patient^{1} + Benefactive^{2}	throwing	throwing
3	L	Agent^{1} + Action^{1} + Patient$^{2(A)}$ + Benefactive$^{2(B)}$	giving	throwing

1- Contrast between Input & Output; 2- Variation within Input and Output; 2(A)- Variation within Input only; 2(B)- Variation within Input only

Appendix 4

Results of regression analyses for NRDLS production and comprehension scores unadjusted and adjusted for potential confounds, data wave and delayed intervention

	T1 - T2				T2 - T3				T1 - T3			
	B	LL	UL	Sig	*B*	LL	UL	Sig	*B*	LL	UL	*p*
NRDLS Comprehension SS												
Treatment arm	-0.83	-5.96	4.30	0.748	6.01	1.68	10.34	**0.007**	3.60	-1.99	9.19	0.205
Adjusted for wave												
Arm group	-0.92	-5.75	3.92	0.708	6.07	1.76	10.39	**0.006**	3.36	-1.88	8.61	0.206
Wave	8.36	3.81	12.91	**0.000**	2.85	-1.66	7.37	0.213	9.49	4.58	14.40	**0.000**
Adjusted for Delayed Intervention												
Arm group	-0.83	-5.99	4.34	0.751	5.80	1.54	10.07	**0.008**	3.71	-1.88	9.31	0.191
DI	0.15	-6.47	6.77	0.964	5.57	0.04	11.10	**0.048**	3.79	-3.42	10.99	0.299
NRDLS Production SS												
Arm group	3.15	-1.11	7.41	0.145	3.57	-0.14	7.29	0.059	6.24	1.74	10.73	**0.007**
Adjusted for Wave												
Arm group	3.26	-1.02	7.53	0.134	4.03	0.47	7.58	**0.027**	6.12	1.81	10.43	**0.006**
Wave	1.82	-2.59	6.23	0.415	5.80	2.20	9.40	**0.002**	6.98	2.61	11.34	**0.002**
Adjusted for delayed intervention												
Arm group	3.17	-1.11	7.45	0.145	3.54	-0.12	7.20	0.058	6.04	1.56	10.52	**0.009**
Delayed Intervention	-0.86	-6.62	4.89	0.767	4.73	0.05	9.41	**0.048**	4.59	-1.53	10.71	0.140

Key: NRDLS: New Reynell Developmental Language Scales; SS: Standard Score;

Results of Bootstrapped regression analyses for NRDLS production and comprehension scores unadjusted and adjusted for potential confounds, data wave and delayed intervention

	T1 - T2				T2 - T3				T1 - T3			
	B	LL	UL	Sig	*B*	LL	UL	Sig	*B*	LL	UL	*p*
NRDLS Comprehension SS												
Arm group	-0.83	-5.92	4.97	0.744	6.01	1.76	10.18	**0.009**	3.60	-1.98	9.07	0.212
Adjusted for wave												
Arm group	-0.92	-5.78	3.90	0.695	6.07	1.97	10.50	**0.005**	3.36	-2.47	8.76	0.242
Wave	8.36	4.37	12.96	**0.001**	2.85	-1.50	7.11	0.185	9.49	4.86	14.23	**0.001**
Adjusted for Delayed Intervention												
Arm group	-0.83	-5.85	4.58	0.745	5.80	1.70	10.09	**0.007**	3.71	-2.03	9.61	0.227
DI	0.15	-5.98	6.12	0.961	5.57	0.71	10.84	**0.030**	3.79	-2.85	10.65	0.246
NRDLS Production SS												
Arm group	3.15	-0.95	7.27	0.134	3.57	-0.34	7.61	0.079	6.24	1.71	10.91	**0.016**
Adjusted for Wave												
Arm group	3.26	-1.22	7.49	0.155	4.03	0.75	7.97	**0.027**	6.12	1.75	10.49	**0.010**
Wave	1.82	-2.29	5.90	0.396	5.80	2.12	9.94	**0.005**	6.98	2.72	11.11	**0.003**
Adjusted for Delayed Intervention												
Arm group	3.17	-1.33	7.57	0.170	3.54	-0.30	7.16	0.073	6.04	1.33	10.47	**0.014**
Delayed intervention	-0.86	-7.48	6.01	0.794	4.73	0.47	8.46	0.022	4.59	-2.39	11.97	0.218

Key: NRDLS: New Reynell Developmental Language Scales; SS: Standard Score;

Appendix 5. Effect Size interpretation guidance recommended by the Education Endowment Foundation (Coe et al 2013)

	Effect size		Description
Months progress	from	to	
0	-0.01	11	Very low or no effect
1	0.02	0.09	Low
2	0.01	0.18	Low
3	0.19	0.26	Moderate
4	0.27	0.35	Moderate
5	0.36	0.44	Moderate
6	0.45	0.52	High
7	0.53	0.61	High
8	0.62	0.69	High
9	0.7	0.78	Very high
10	0.79	0.87	Very high
11	0.88	0.95	Very high
12	0.96	>1.0	Very high