CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	The effectiveness of individualized morphosyntactic target identification and explicit intervention using the SHAPE CODING system for children with developmental language disorder and the impact of within-session dosage
Authors	Ebbels, Susan H.;Gadd, Mollie;Nicoll, Hilary;Hughes, Lucy;Dawson, Nicola;Burke, Caroline;Calder, Samuel D.;Frizelle, Pauline
Publication date	2024-06-19
Original Citation	Ebbels, S.H., Gadd, M., Nicoll, H., Hughes, L., Dawson, N., Burke, C., Calder, S.D. and Frizelle, P. (2024) 'The effectiveness of individualized morphosyntactic target identification and explicit intervention using the shape coding system for children with developmental language disorder and the impact of within-session dosage', Language, Speech, and Hearing Services in Schools, 55(3), pp. 803–837. Available at: https://doi.org/10.1044/2024_LSHSS-23-00098.
Type of publication	Article (peer-reviewed)
Link to publisher's version	10.1044/2024_LSHSS-23-00098
Rights	© 2024. This is the accepted manuscipt version of an article published in Language, Speech, and Hearing Services in Schools is made available under a under a Creative Commons Attribution 4.0 International license. The version of record is available at https://doi.org/10.1044/2024_AJSLP-23-00375 - https://doi.org/10.1044/2024_AJSLP-23-00375, https://creativecommons.org/licenses/by/4.0/
Download date	2026-11-17 04:50:45
Item downloaded from	https://hdl.handle.net/10468/16104

University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

The effectiveness of individualised morphosyntactic target identification and explicit intervention using the SHAPE CODING™ system for children with Developmental Language Disorder (DLD)

Susan H Ebbels[1,2], **Mollie Gadd**[1], **Hilary Nicoll**[1], **Lucy Hughes** [1,2], **Nicola Dawson**[1,5], **Caroline Burke**[1], **Samuel D Calder**[3], **Pauline Frizelle**[4]

[1]Moor House Research and Training Institute, Moor House School & College, Oxted, UK

[2] Division of Psychology and Language, University College London, UK

[3] Health Sciences, College of Health and Medicine, University of Tasmania, Launceston, Australia

[4] Department of Speech and Hearing Sciences, University College Cork, Republic of Ireland

[5] Department of Experimental Psychology, University of Oxford, UK

Correspondence to: Susan Ebbels, Moor House Research and Training Institute, Moor House School & College, Mill Lane, Hurst Green, Oxted, Surrey, RH8 9HN, UK, ebbelss@moorhouseschool.co.uk

Conflict of Interest: Susan Ebbels, Hilary Nicoll, Caroline Burke, Lucy Hughes and Nicola Dawson are (and Mollie Gadd was at the time of the project) employed by Moor House School & College which sells resources and courses on the SHAPE CODING™ system. There are no other conflicts of interest.

Ethics: This research was approved by the Moor House Ethics Committee

ABSTRACT

Purpose

To investigate the effectiveness of a highly individualised morphosyntactic intervention using the SHAPE CODING™ system and to investigate whether progress with intervention and maintenance of any progress differs between participants.

Method

Eight children with DLD aged 8;0-10;10 received individualised intervention for morphosyntax for 30 minutes 1:1 once per week for 20 weeks (10 hours). In a multiple baseline design, following at least four baseline probe tests, two grammatical targets per session received explicit instruction until they reached criterion (90%), when the next target was introduced. Maintenance testing of completed targets was also carried out.

Results

In total, across the eight participants, 46 morphosyntactic targets (from nine broad target groups) received intervention, with 4-8 intervention targets per participant. These included 27 unique targets, highlighting the individualised outcomes of the target identification process.

Scores on probe tests during and after intervention were significantly higher than during the baseline phase. Taking average scores across the eight participants, while there was slight progress during the baseline phase, the slope during the intervention phase was much steeper and highly significant. Progress was maintained. Two of the eight participants (who had received similar intervention previously) showed significantly faster progress with intervention while one (with the lowest attention score) did not make progress.

Conclusions

With the exception of the participant with only moderate attention levels, the individualised intervention was highly effective and efficient. Thus, the individualised target identification process and intervention method and techniques merit further research in a larger group of children and should be considered for use with similar children with DLD with high attention levels, but with grammatical difficulties.

Introduction

Language is at the heart of children's learning, social interaction and wider development, yet approximately 10% of children have a language disorder, with 7% meeting the criteria for Developmental Language Disorder (DLD) (Norbury, Gooch, Wray, Baird, Charman, Simonoff et al., 2016). In an increasingly information-based economy, difficulties understanding and using spoken and written language put children at risk of poor educational (Nippold, 2010), employment (Conti-Ramsden, Durkin, Toseeb, Botting, & Pickles, 2018) and mental health outcomes (Yew & O'Kearney, 2013). It is therefore unsurprising that the top research priorities identified by researchers, practitioners and individuals with lived experience of DLD all focus on establishing effective interventions (Kulkarni, Chadd, Lambert, Earl, Longhurst, McKean et al., 2022), with the second highest priority being identification of the "specific characteristics of evidence-based DLD interventions which facilitate progress towards the goals of an individual with DLD" (p.958).

In typical language development, children extract the statistical regularities of the language(s) they are exposed to without effort or awareness. However, this form of implicit learning is challenging for children with DLD (Lammertink, Boersma, Wijnen, & Rispens, 2017), leading to significant and persistent difficulties in forming accurate and complex sentences to express their thoughts. Indeed, while children with DLD have deficits across language domains, their grammatical abilities are particularly affected (Leonard, 2014), such that they often produce short, simple sentences (Nippold, Mansfield, Billow, & Tomblin, 2009) containing grammatical errors (Scott & Windsor, 2000) and have difficulty understanding and using complex sentences (Frizelle & Fletcher, 2014a, 2014b), all of which affect their ability to communicate with others and access the educational curriculum.

Intervention approaches for grammar employ either predominantly implicit or explicit methods (see Ebbels, 2014), or a combination. Predominantly implicit interventions make target

forms more frequent and salient and have been shown to improve specific aspects of expressive syntax in young pre-school or early school-aged children with primarily expressive language deficits or delays, when between 15 and 60 hours of intervention is delivered 1:1 by an SLT or parent (Fey, Cleave, Long, & Hughes, 1993; Fey, Cleave, & Long, 1997; Owen van Horne, Fey, & Curran, 2017; Plante, Ogilvie, Vance, Aguilar, Dailey, Meyers et al., 2014). Few studies of implicit intervention include older children or those with more wide-ranging difficulties including receptive language.

Explicit intervention approaches tell the children the grammatical rules of their language, often using visual cues in order to capitalise on the relative strengths in declarative memory and visuo-spatial skills of children with DLD (Lum, Conti-Ramsden, Page, & Ullman, 2012). Of the explicit intervention approaches for morphosyntax, the SHAPE CODING™ system has the strongest evidence base. RCTs conducted in specialist settings demonstrate improvements in targeted aspects of grammar (and generalisation to non-targeted items following the same rule) in school-aged children (aged 5-16 years) including those with receptive language deficits, when intervention is delivered for 5-10 hours 1:1 by an SLT (Calder, Claessen, Ebbels, & Leitão, 2021; Ebbels, Maric, Murphy, & Turner, 2014; Ebbels, van der Lely, & Dockrell, 2007). Given the shorter amount of time in intervention compared with implicit approaches, it seems that explicit approaches may be more efficient (at least for children with DLD over the age of 5-6 years), even though both approaches are effective. Indeed Frizelle, Tolonen, Tulip, Murphy, Saldana, and McKean (2021) drew a similar conclusion in their review.

Many intervention studies targeting morphosyntax in children with DLD consider a single target of intervention (e.g., past tense, Calder et al., 2021; causal adverbials, Curran & Owen van Horne, 2019; past tense, Owen van Horne et al., 2017) while others choose particular targets from a small range of targets (e.g., coordinating conjunctions, Ebbels et al., 2014; verb morphology, Fey, Leonard, Bredin-Oja, & Deevy, 2017; Plante, Mettler, Tucci, & Vance, 2019).

However, different children with DLD may have difficulties with a wide range of grammatical structures, both syntactic and morphological in comprehension and/or expression and clinicians need to be able to take an individualised approach to the identification and prioritisation of targets for intervention as well as potentially provide intervention for more than one target at a time.

In order to maximise the efficiency of intervention and to avoid diminishing returns (Frizelle et al., 2021), clinicians need a system for monitoring progress with intervention. As rates of progress are likely to vary between children and targets, it is important to regularly monitor progress against pre-specified criteria to identify when a target has been achieved (and therefore when intervention on that target can cease) or when a child's performance has plateaued (and they therefore perhaps needs to move onto a different target).

In this study, we aimed to maximise efficiency by personalising intervention to each individual child, such that they worked on individualised targets, with regular monitoring of progress and introduction of a new target as soon as a previous target had either been achieved or plateaued.. We also aimed to evaluate the effectiveness of an explicit intervention focusing on these individualised targets and consider whether there were any differences in progress with intervention between children and targets and whether any progress generalised to other general language measures (mean length of utterance, errors per utterance, clausal density) and standardised language tests. Our specific research questions were:

1. Is there an overall effect of intervention?
2. Does the number of days in the baseline, intervention and/or maintenance phases predict scores and do any of the individual participants show differing patterns or trends?
3. Does the cumulative number of teaching episodes predict outcomes and do any individual participants show differing patterns or trends?
4. Do any specific targets show differing patterns or trends?

5. Do participants show any changes between the start and end of the study on standardised language tests or other grammatical measures (mean length of utterance, complexity, or errors per utterance)?

In our companion paper (REF), we also aimed to address questions around within session dosage and its potential impact on progress with intervention.

Method

Study Design

Our study was a single blind within-participant multiple baseline design consisting of four stages: 1) pre-intervention target identification where individualised targets were identified for each participant, 2) pre-intervention baseline measures, 3) intervention using the SHAPE CODING system with weekly probe tests to measure progress towards pre-defined stopping criteria, and 4) a maintenance phase during which retention probes were planned for administration for each completed target, 2, 6 and 14 weeks post-intervention.

Ethics

The research was approved by the [redacted] Ethics Committee

Participants

This intervention study was carried out at a specialist day and residential school for children with (developmental) language disorders in the UK. All children in the school have language skills which significantly affect their learning. The majority meet the criteria for DLD but some children with associated biomedical conditions, but where language is judged to be the primary barrier to learning, also attend the school. This study focused on children in Key Stage 2, aged between 7 and 11 years. Eight children (6 boys and 2 girls) from two classes were judged to be suitable for the project, i.e., morphosyntactic difficulties were a priority area for intervention and their parents consented to participation (no parents refused consent). Their ages ranged from 8;0-10:10 at the start of the intervention. All eight participants had a diagnosis of DLD, with seven also having a diagnosis of speech sound disorder, one of whom additionally had an ADHD

diagnosis. All eight participants had English as a main language. One had limited exposure to another language, but had attended an English-speaking nursery and previous school setting. We did not index the socio-economic status of our participants as due to the nature of the ongoing relationships between parents of participants and interventionists, we deemed it unnecessarily intrusive to ask parents for information on levels of education or income. This information is also perhaps less relevant in a within participant design.

To profile the participants' language abilities prior to the intervention commencing and to inform the identification of appropriate treatment targets, the following language assessments were completed 1) the Structured Photographic Expressive Language Test-Third Ed (SPELT-3: Dawson et al., 2005), 2) the narrative subtest from the Assessment of Comprehension and Expression (ACE, Adams, Cooke, Crutchley, Hesketh, & Reeves, 2001) and 3) the 'Beach' Story subtest from the Expression, Reception, and Recall of Narrative Instrument (ERRNI; Bishop, 2004). The 'Beach' story was chosen, as it was found to elicit more complex sentences from typically developing children than the alternative 'Fish' Story (Frizelle, Thompson, McDonald, & Bishop, 2018). In addition, we also carried out the Test for the Reception of Grammar- Electronic (TROG-E; Bishop, 2003) to measure children's understanding of syntax and monitor any change with this following intervention. Summaries of these scores are shown in Table 1, with individual scores in the Supplementary Information 1-3. These scores show that all participants had difficulties with both expressive and receptive grammar.

Table 1: Mean raw and standard scores on standardised assessments pre-intervention

Assessment	Raw		Standard/Scaled	
	Mean	SD	Mean	SD
SPELT	28.12	10.47	60.57[a]	21.62
TROG	7.12	3.31	65.00 [a]	10.92
ACE grammar	6.25	2.38	3.38 [b]	0.74

Assessment	Raw		Standard/Scaled	
	Mean	SD	Mean	SD
ACE information	8.88	4.55	6.00 [b]	2.88
ERRNI MLU	5.51	1.20	69.25 [a]	8.71
ERRNI content	8.62	2.50	70.38 [a]	4.60

[a] = standard score (1SD range 85-115) [b] = scaled score (1SD range: 7-13)

General Treatment context

To reflect the usual protocol within the school, treatment was administered in termly blocks (roughly 10 weeks) in the second (January - April 2022) and third (April - July 2022) terms of the school year. Intervention was delivered by the second author, a speech and language therapist working in the school, who had 3 years' experience using the SHAPE CODING system and had attended the Part 1 and Part 2 SHAPE CODING courses. The treating SLT and first author (who developed the SHAPE CODING system) initially met weekly face-to-face to discuss progress against targets for all participants and the next steps for the following session. Later in the project, communication was predominantly via electronic messages and shared spreadsheets. In cases of absence, intervention sessions were covered by the first author. Over the whole study, the first author provided eight intervention sessions in total across six different participants (once or twice each).

With marginal variation, participants received intervention on the same day and at a similar time each week, however, some timetable variation was required so that sessions did not always impact the same class lessons. Teachers were blind to which specific targets had been assigned to each participant and were asked to support language in a more general way in the classroom, rather than focussing on specific morphosyntactic targets. The participants did not receive any additional intervention targeting expressive grammar during the project period.

Intervention Target Identification

Following transcription of the language samples gathered from the assessment battery previously described, the samples were analysed against a list of 133 potential morphosyntactic

targets split into 11 target groups: main clauses, questions, conjoining, adverbials, complement clauses, passives, relative clauses, tense and aspect, subject verb agreement, noun phrases and negatives. Each target has a code which indicates its target group and its relative position within that target group (e.g., NP9 is the 9th structure/target in the Noun Phrase target group). Performance for each child on the language samples was recorded on a copy of the spreadsheet developed by the first and third authors (an updated version following changes due to learning from this project is available here: https://shapecoding.com/resources/grammar-spreadsheet-free/). The first stage of the target identification progress identifies potential targets for further testing on probe tests. The spreadsheet includes formulae that indicate whether a target form needs further probe testing. If a target form is not used correctly at least twice, or if it is attempted but this results in errors for more than 25% of the production attempts, it is considered a potential target that requires further testing using a probe test (see below).

Both the first and second authors filled out the spreadsheet for three of the participants and their agreement on whether a potential target required a probe test or not was 92%. The five other participants were each scored by just one of these two authors. Most disagreements were either because one author had missed a child production matching a target form or because when the child had produced an error, the authors had not always agreed on which target to mark the error. For example, for "then a parrot came and chattering to the monkey", one author had marked the error on the past tense target and the other on both past tense and verb phrase coordination. As the use of verb phrase coordination was in fact accurate, the subsequent agreed decision was to credit the participant with this. Identified differences in marking were resolved through discussion and scores for other participants amended as appropriate before the probe testing started.

Probe testing

In the second stage of the target identification process, probe tests are carried out for those potential targets with the highest priority in the suggested hierarchy on the target identification spreadsheet.

The third, sixth and first authors (listed in order of contribution levels) created probe tests for all potential targets consisting of 10 stimulus items designed to elicit each target form plus one model and one practice item. Where a participant's production accuracy fell below 90% on the probe test, that target was identified for intervention. However, if performance rose above 90% during the baseline phase, the target was dropped.

Probes were administered once a week during the baseline and intervention phases and less frequently during the maintenance phase by the sixth author who was blind to whether each target for each participant was in the baseline, intervention or maintenance phase. Probe sessions were always on a different day from intervention sessions. In cases of absence of a participant or the testing SLT on their test day, the first author carried out the testing (never on the same day as she delivered any intervention). The first author carried out tests on 24 days (often to catch up a single participant) and the sixth author on 66 days (usually with multiple participants). The probe tests aimed to establish the participants' baseline ability and then any change with intervention in their ability to generalise intervention targets to untreated contexts (both linguistic contexts and with a different adult), using lexical items (where possible) that were not used in intervention (for more information see our companion paper, REF).

All probe tests were scored by the sixth and first authors with regular discussions and recording of scoring criteria to ensure consistent scoring between the two scorers. The main issues were around whether to score a particular utterance as incorrect versus not applicable,

particularly when the participant had not attempted a particular construction. Decisions around this were added to scoresheets to ensure consistency[1].

Pre-treatment baseline phase

Following the identification of intervention targets, four weekly baseline measurements were administered for the top four priority targets for each participant. Intervention then started for the first two targets and baseline measures continued for the next two targets. Each time a participant reached a stopping criterion for a target (see "stopping criteria" section for more information), intervention ceased for that target, a new target was introduced into intervention, and baseline testing began for the next target in the hierarchy. The aim was to always have at least four baseline test points for each target before intervention began on that target. In most cases, the baseline was longer, but its length varied according to the speed with which each participant reached the stopping criterion for the previous target. In a couple of cases, a new target was needed before the next target had a baseline with at least four timepoints, in this case, intervention was only provided on one target until the next target had at least four baseline timepoints.

Intervention protocol

All participants received intervention using the SHAPE CODING system, which provides a child-friendly explanation of the rules behind each target structure supported with visual templates. The rules and SHAPE CODING templates for all structures included in this project are in Appendix A. Broad intervention steps for each target were developed by the third and first authors and made available to the treating SLT (second author). The precise techniques, procedures and feedback hierarchy used across all targets are described in our companion paper (REF). These were combined into teaching episodes which provided the minimum amount of

[1] Examples of answers coded as "not applicable": 1) the target was plural auxiliary *are/were*, but the participant used a singular subject, so a plural verb not required, or the auxiliary was missing entirely - these would both be coded as an error on a different target, 2) the target was use of adverbial clauses, but the participant failed to use the target subordinating conjunction, 3) the target was formation of "wh" questions, but the participant did not attempt to ask a question (perhaps misunderstanding the task).

support necessary for the child to produce the target correctly, at which point the teaching episode was deemed to be complete. Thus, the number of completed teaching episodes per target per session was equivalent to the number of correct child productions of a target in a session.

Participants received intervention individually in a room within the school. Each participant worked on two intervention targets per session. The order of focus on the two targets alternated from session to session. For some participants both targets were morphological, for others both were syntactic and for others the targets were mixed between morphology and syntax.

Dosage

The number of teaching episodes per target per session (10, 20 or 30) was manipulated. and the method and effects of this are reported in our companion paper (REF). Here, the number of teaching episodes per target per session is included only as a random variable. However, we use cumulative teaching episodes (the total number of teaching episodes to date on a target) as a predictor of progress in the analyses addressing research questions 3 and 4.

Stopping criteria

We implemented two stopping criteria throughout the intervention. The first criterion (*target achieved*) was to identify when participants had made sufficient progress on a given target that they no longer required intervention on that target. 90% production accuracy was deemed an appropriate indication of this.

The second criterion (*discontinue target*) was for those participants who were not, or no longer, making progress on a target. There are obvious ethical issues in continuing to focus on a target with little or no (continued) progress when a child could benefit from input on a different target. The minimum dosage before the discontinue target rule was applied was 180 teaching episodes. This value was based on the average number of teaching episodes required to observe progress (and needed to be divisible by 10, 20 and 30 to accommodate the three possible dosages). For each participant and target, we were then able to run a linear regression to examine whether the proportion correct on that target changed significantly as a function of cumulative teaching

episodes (once the minimum of 180 teaching episodes had been reached). If the outcome was non-significant, intervention was discontinued for that target.

Intervention fidelity and child attention

The number of sessions received per participant ranged from 16 to 20 sessions out of the 20 planned sessions (see Table 2). Missed sessions were mainly due to high levels of absence as a result of sickness during 2022 among both adults and students in UK schools. All sessions were video recorded and 20% were randomly selected for analysis by the fourth author, who examined adherence to intervention procedures (see our companion paper, REF, for further information) and children's attention within sessions.

Children's attention in sessions was considered to be a potential moderating factor, which could affect the delivery and outcomes of intervention (Carroll et al., 2007). In order to explore this, we used the Attention subscale of the Pivotal Behaviour Rating Scale (Revised; Mahoney, 1998), which rates children's attention in four different areas: (1) their general ability to focus on an activity; (2) their level of persistence within tasks; (3) their active involvement in the session and (4) their compliance with adult suggestions. Each area is rated out of five, where 1=very low, 2=low, 3=moderate, 4=high and 5=very high and an average score for attention is calculated.

Mean attention scores for individual participants across analysed sessions are shown in Table 2. On average, their scores were high, but there was variability within the sample.

Table 2 – Intervention sessions received and mean attention scores in rated video sessions

	Participant								
	ID1	ID2	ID3	ID4	ID5	ID6	ID7	ID8	Mean
Total number of intervention sessions received	16	17	19	20	17	20	20	17	18.25
Mean attention score across videoed sessions	4.92	4.40	4.44	4.92	3.96	3.13	5	4	4.34

ID6 showed the lowest average attention score (but was not the child with the ADHD diagnosis). He scored in the "moderate" range, indicating that there were sustained periods during which he appeared focused on what he was doing and other extended periods in which he seemed disinterested in the activities (Mahoney, 1998). In contrast, ID7's attention was rated as "very high" across all activities and sessions, reflecting that this participant was highly involved throughout the observed sessions. ID5's scores fell between the "moderate" and "high" range. All remaining participants were rated as showing a "high" to "very high" level of engagement, staying focused on the activities and appearing to derive satisfaction from their participation.

Maintenance phase

During the maintenance phase, we aimed for retention probes to be administered 2, 6 and 14 weeks after intervention ceased. However, the study needed to finish at the end of the school year so there was insufficient time for the targets completed later in the study to have such a long maintenance phase. We therefore collected as much data as we could at the end of the study, aiming to get as many maintenance retention tests carried out as possible. This meant that the time between maintenance tests was greatly reduced for some targets. Due to the variation in lengths of time between maintenance tests, the number of days in maintenance was used as the main predictor for any analyses, so that we could see if scores changed with time after intervention ceased. At the end of the project, some targets had not yet reached the stopping criteria so were still in the intervention phase and therefore had no maintenance data.

Analysis

All data were analysed using logistic regression. This type of general linear model (Howell, 2010) is used for a binomially distributed dependent variable (i.e., correct versus incorrect), and accounts for random effects, such as differences between individuals, dosage or

targets. Analyses were carried out in R version 4.3.0 using the glmmTMB package (version 1.1.7).

Results

Research Question 1 – Is there an overall effect of intervention?

In order to establish whether there was an overall effect of intervention, we first ran a mixed effects model with the fixed effect of phase (baseline, intervention or maintenance), ignoring time within each phase, but including random intercepts of participant, dosage group and target (this was the random effects structure that explained the most variance and converged). These data are shown in the violin and box plot in Figure 1. This model showed a significant overall effect of phase where relative to the baseline phase, the odds of a correct response were significantly higher in the intervention, *Odds Ratio*: 4.49 (*CI*: 3.86-5.22), $p<.001$ and maintenance phases, *OR*: 6.19 (*CI*: 5.13-7.47), $p<.001$, thus indicating the general effectiveness of the intervention.

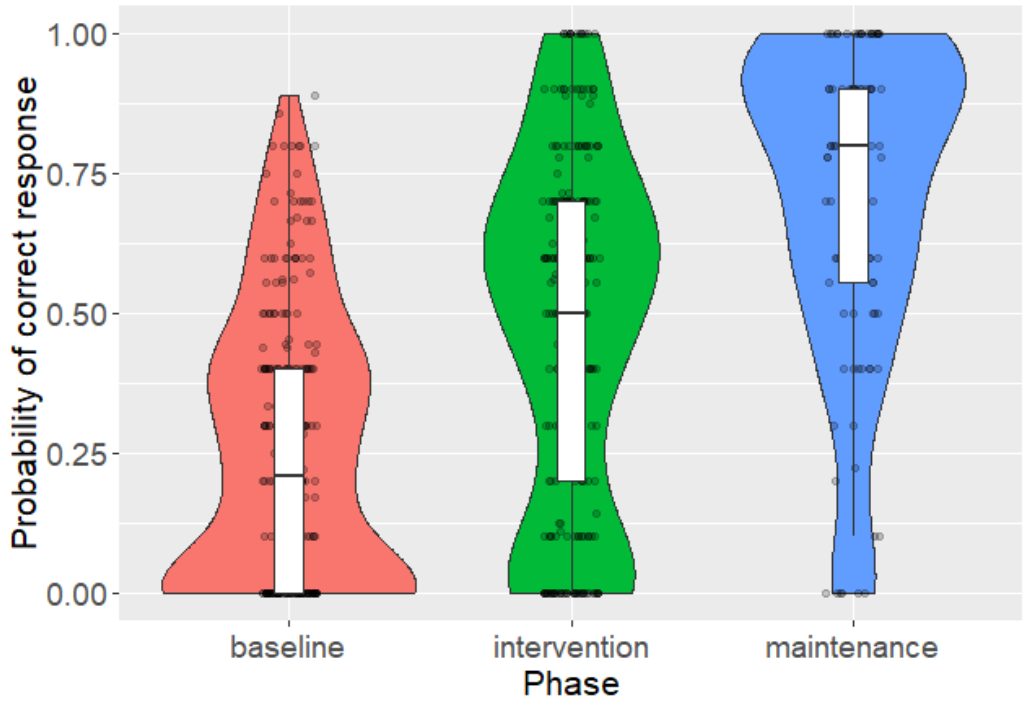


Figure 1 – Violin and box plot showing the probability of a correct response during each of the phases

Research Question 2 – Does the number of days in the baseline, intervention and/or maintenance phases predict scores and do any individual participants show differing patterns or trends?

The above analysis ignored the effect of time within phases and included participants as a random factor. This means that the performance of individual participants could not be evaluated and also, using the mean performance in each phase masked any change in performance over time within a phase, thus potentially accounting for the higher scores in the maintenance phase than in the intervention phase. In the following analysis, we specifically consider the fixed effects of participant and days within each phase and the interaction between them, to establish whether all eight participants responded to the intervention in a similar manner. Thus, we included two fixed effects: participant (sum-coded) and days in phase, and their interaction. In this analysis, we were not interested in the effect of the target (which will be considered in the next analysis), or the dosage group (which is considered in our companion paper, REF). Thus, dosage group, target and the interactions of these with days in phase were included as potential random variables.

In order to analyse each phase in detail, we considered each of the three phases separately. This had the advantage that some data points could appear in more than one phase: i.e., the last data point of the baseline phase was included as the intercept for the intervention phase and the last data point of the intervention phase was included as the intercept for the maintenance phase. For the intervention and maintenance phases, the best model in terms of random variables included intercept for dosage group and by-target slopes with days in phase. For the baseline phase, the by-target slope with days in phase was no better than the intercept for target. However, for consistency, the baseline phase was analysed with the same random factors as the intervention and maintenance phases. The full model which included the fixed effects of days in phase, participant and their interaction, improved the model fit for all phases. The results are shown in Table 3.

Table 3: Model for each phase separately with fixed factors of days in phase and participant

	Baseline Phase			**Intervention Phase**			**Maintenance Phase**		
Predictors	*Odds Ratios*	*95% Confidence Interval*	*p*-value	*Odds Ratios*	*95% Confidence Interval*	*p*-value	*Odds Ratios*	*95% Confidence Interval*	*p*-value
Days in Phase	1.014	1.003 – 1.024	**0.010**	1.083	1.033 – 1.134	**0.001**	0.993	0.985 – 1.002	0.132
Days in Phase * ID1	1.017	0.992 – 1.043	0.188	1.010	0.992 – 1.030	0.273	1.010	0.994 – 1.026	0.208
Days in Phase * ID2	0.983	0.958 – 1.010	0.209	1.035	1.015 – 1.056	**0.001**	1.004	0.988 – 1.020	0.616
Days in Phase * ID3	1.002	0.982 – 1.023	0.822	0.978	0.924 – 1.035	0.434	1.025	0.986 – 1.067	0.214
Days in Phase * ID4	0.986	0.968 – 1.004	0.123	1.028	1.002 – 1.054	**0.034**	0.993	0.976 – 1.011	0.459
Days in Phase * ID5	1.019	0.997 – 1.041	0.085	0.997	0.983 – 1.012	0.698	0.983	0.962 – 1.003	0.102
Days in Phase * ID6	0.994	0.970 – 1.020	0.670	0.926	0.901 – 0.951	**<0.001**	1.014	0.997 – 1.032	0.106
Days in Phase * ID7	0.995	0.980 – 1.010	0.517	1.015	1.002 – 1.028	**0.029**	0.986	0.974 – 0.999	**0.037**
Days in Phase * ID8	1.004	0.985 – 1.024	0.681	1.016	0.102 – 1.030	**0.029**	0.984	0.969 – 0.999	**0.042**
Random Effects									
σ^2	3.29			3.29			3.29		
τ_{00}	0.05 $_{\text{Dosage Group}}$			0.53 $_{\text{Dosage Group}}$			0.24 $_{\text{Dosage Group}}$		
	2.03 $_{\text{Target}}$			2.27 $_{\text{Target}}$			3.56 $_{\text{Target}}$		
τ_{11}	0.00 $_{\text{Target. Days in Phase}}$			0.01 $_{\text{Target. Days in Phase}}$			0.00 $_{\text{Target. Days in Phase}}$		
ρ_{01}	-0.25 $_{\text{Target}}$			0.11 $_{\text{Target}}$			0.01 $_{\text{Target}}$		
ICC	0.41			0.87			0.56		
N	3 $_{\text{Dosage Group}}$			3 $_{\text{Dosage Group}}$			3 $_{\text{Dosage Group}}$		
	26 $_{\text{Target}}$			26 $_{\text{Target}}$			21 $_{\text{Target}}$		
Observations	222			301			131		
Marginal R^2 / Conditional R^2	0.088 / 0.464			0.235 / 0.897			0.168 / 0.636		

Table 3 shows that during the baseline phase there was a marginally significant increase with days in phase across all participants, with no significant difference in slope for the different participants. During the intervention phase, there was a highly significant increase in scores on target probes with days in intervention in the grand mean across all participants, which was steeper than the slope during the baseline phase. However, five participants differed significantly from the grand mean – four made significantly greater progress (ID2, ID4, ID7 and ID8) and one made significantly less progress (ID6 – this is investigated further below). During the maintenance phase there was no significant change in the grand mean. Thus, progress was maintained. Two participants showed a marginally significantly different slope to the others during the maintenance phase (ID7 and ID8 – this is further investigated below).

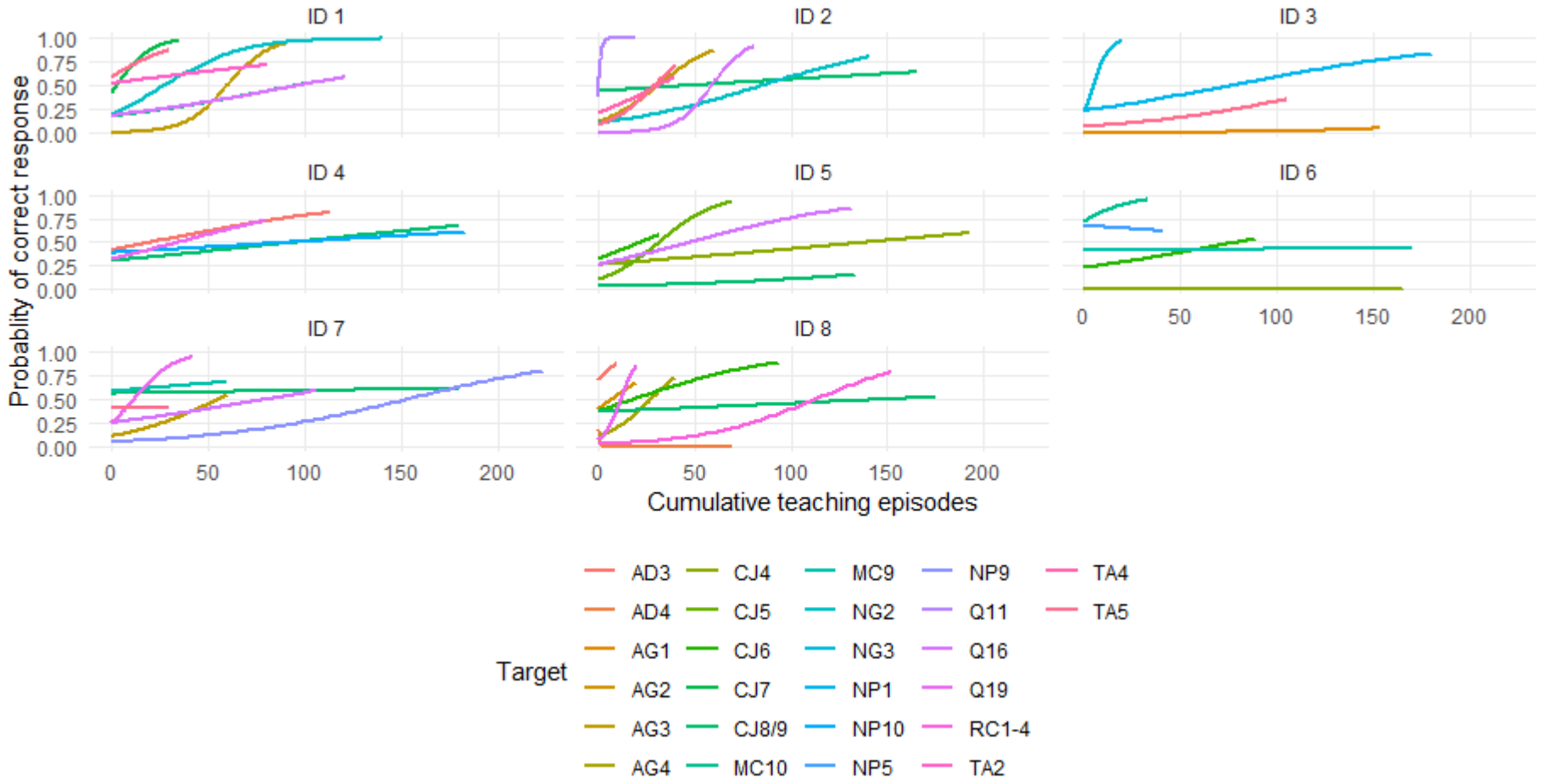
Given that ID 6 had a different slope during the intervention phase from the other participants, we then carried out an analysis with just this participant in the intervention phase. A mixed model analysis included a fixed factor of days since the start of intervention and random intercepts for dosage and target (the model did not converge with by-target slope as a random variable). This analysis showed no significant increase with time during the intervention phase for participant ID6, OR = 0.99 (CI=0.98-1.01), p= .34. Thus, the intervention does not seem to have been effective for this participant.

IDs 7 and 8 showed a different pattern from the others during the maintenance phase. Therefore, we analysed just their scores during this phase. A mixed model analysis included a fixed factor of days since the start of maintenance and random intercepts for dosage group and target (models with random slopes did not converge). This analysis showed a significant decline with time during the maintenance phase for both participants ID7, OR = 0.98 (CI=0.97-0.99), p= .001 and ID8, OR = 0.98 (CI=0.97-0.99), p< .001. Thus, for these two participants, while the intervention appears to have been very effective (they showed faster

progress than the grand mean for all participants), its effects without further input diminished during the maintenance phase.

Research Question 3 – Does the cumulative number of teaching episodes predict outcomes and do any individual participants show differing patterns or trends?

Dosage was manipulated within the study to try to establish whether differing numbers of teaching episodes per session affect the effectiveness or efficiency of the intervention (this is the focus of our companion paper, REF). However, the differing dosage (10, 20 or 30 teaching episodes per session) means that after a given number of intervention sessions, some targets had received three times as many teaching episodes as others. Indeed, our companion paper (REF) showed differences between the dosages in the rate of progress when predicted by cumulative number of intervention sessions, but these differences disappeared when cumulative teaching episodes were instead used as the predictor. Therefore, our next analysis uses cumulative teaching episodes; this is the total number of teaching episodes delivered by each testing point. This predictor also has the advantage of including the baseline, intervention and maintenance phases in the same analysis; all of the testing points in the baseline phase have zero teaching episodes, during the intervention phase the cumulative number of teaching episodes increases and during the maintenance phase the cumulative number of teaching episodes stays constant at the number reached when intervention ceased (as participants receive no more teaching episodes during the maintenance phase).



435

436 *Figure 2 – Probability of correct response by cumulative teaching episodes on individual targets (separate lines for each target, see Appendix A*

437 *for explanations of target codes, more details and examples)*

Figure 2 shows the probability of a correct response for each participant and target with cumulative teaching episodes on the x-axis. Each *teaching episode* included a range of techniques in a specified order, providing the minimum support required for a correct child production. As each *teaching episode* ended with a correct child production, we counted how many times in total the child had correctly produced the target during intervention sessions to date[2] as a proxy for the number of *cumulative teaching episodes*. Because the targets were all individually assigned and intervention ceased (and a new target was introduced) either when the participant reached 90% correct, or when they showed a sustained lack of progress (see section on stopping criteria), the different participants received intervention on a different number of targets, for differing amounts of time. Figure 2 shows that all participants showed progress on at least one target. However, the rate of progress varied greatly both within and between participants, with some participants showing much faster progress on some targets than others.

In order to establish whether using cumulative teaching episodes as the predictor gave similar results to the analysis in research question 2 using cumulative days in intervention, we ran a mixed effects model with two fixed effects: participant (sum-coded i.e., using the grand mean) and cumulative teaching episodes. In this analysis, we were not interested in the effect of the individual targets (which will be considered later), or the dosage group (which is considered in our companion paper, REF). The best fitting model that converged included the fixed effects of participant and cumulative teaching episodes and their interaction and the random intercept of dosage group and by-target random slope with cumulative teaching episodes. The results are shown in Table 4.

[2] We aimed for child productions to match the assigned dosage. While this was mostly achieved, there were a few variations (see section on fidelity in method of out companion paper, REF). Therefore, for our analyses we used the number of correct productions produced by the child, as recorded by the SLT during sessions, rather than the number which were aimed for.

Table 4 shows very similar results to the intervention phase analysis in research question 2 (Table 3). The odds of a correct answer increased significantly with cumulative teaching episodes (4.0% per teaching episode[3]). IDs 2 and 4 still showed significantly greater progress than the grand mean (with increased odds of a correct response of 4.9% and 4.6% per teaching episode respectively) and ID6 significantly less (only 2.7% increased odds per teaching episode). In this analysis, ID1 also showed steeper progress than the grand mean, but IDs 7 and 8 did not (unlike in research question 2, where they improved faster than the grand mean during the intervention phase, but then showed a decrease in the maintenance phase). This is probably because the analysis with cumulative teaching episodes combines all three phases of the study with all tests after the end of intervention having the same number of teaching episodes, thus ID7 and 8's initially stronger gains with intervention are cancelled out by the loses during the maintenance phase. This analysis together with the analysis in research question 2 leads us to conclude that on average the participants benefited from the intervention, with the exception of ID6. ID2 and ID4 seem to have shown particularly strong progress.

Table 4: Model showing odds ratios with fixed predictors of cumulative teaching episodes and participant (sum-coded)

Predictors	*Odds Ratios*	*95% Confidence Interval*	*p*-value
Cumulative teaching episodes	1.040	1.018 – 1.062	**<0.001**
Cumulative teaching episodes * ID 1	1.007	1.001 – 1.013	**0.020**
Cumulative teaching episodes * ID 2	1.009	1.003 – 1.014	**0.002**
Cumulative teaching episodes * ID 3	0.984	0.967 – 1.000	0.057
Cumulative teaching episodes * ID 4	1.006	1.001 – 1.011	**0.019**
Cumulative teaching episodes * ID 5	1.002	0.997 – 1.007	0.494

[3] This finding matches exactly that in research question 1 of our companion paper (REF) which looked at the main effect of cumulative teaching episodes with dosage group sum-coded as a main predictor and with participants (and target) as a random factor.

Cumulative teaching episodes * ID 6	0.987	0.980 – 0.995	**0.001**
Cumulative teaching episodes * ID 7	1.004	1.000 – 1.008	0.076
Cumulative teaching episodes * ID 8	1.002	0.998 – 1.007	0.320

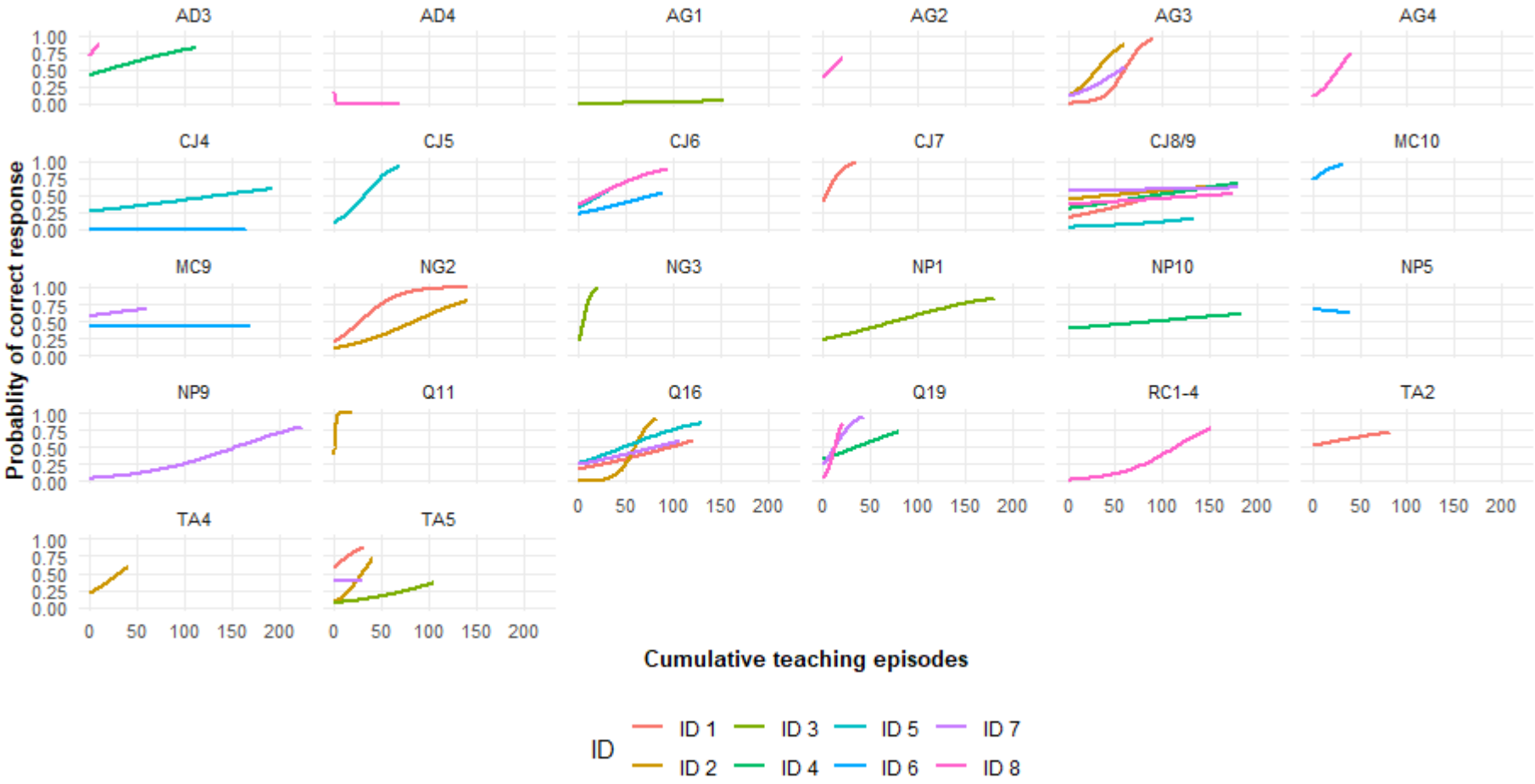
Random Effects

σ^2	3.29
$\tau_{00\ Dosage\ Group}$	0.22
$\tau_{00\ Target}$	1.99
$\tau_{11\ Target.Cumulative\ teaching\ episodes}$	0.00
$\rho_{01\ Target}$	0.29
ICC	0.84
$N_{\ Dosage\ Group}$	3
$N_{\ Target}$	26
Observations	573
Marginal R^2 / Conditional R^2	0.201 / 0.871

Given that ID 6 had a shallower slope than the grand mean across all participants, we then repeated the analysis with just this participant. This showed no significant increase with cumulative teaching episodes for participant ID6, *OR* = 1.01 (*CI*=1.00-1.02), *p*= .17, reinforcing the finding in research question 2 that the intervention seems to have had little effect for this participant.

Research Question 4 – Do any specific targets show differing patterns or trends?

In this analysis, we wanted to consider whether certain targets showed more or less progress than others. Figure 3 shows the data for each of the 27 unique targets with cumulative child teaching episodes on the x-axis. This shows that most participants made progress with most targets, but the rates of progress differed greatly between targets and participants, with some targets showing very rapid progress and others much slower progress. It was not possible to analyse these data statistically as there were many targets and several were only a target of intervention for a single participant.



493

494 *Figure 3: Probability of a correct response by participant and target (see Appendix A for descriptions of each target and examples of each*

495 *target).*

We considered which targets were discontinued due to limited progress. Table 5 shows how many targets were 1) achieved, 2) still in progress at the end of the study, or 3) discontinued due to limited progress.

Table 5: Number of targets achieved, discontinued or in progress at the end of the study by participant

	Participant								
	ID1	ID2	ID3	ID4	ID5	ID6	ID7	ID8	Mean
Targets achieved (90% correct)	5	5	1	2	3	2	3	5	3.25
Targets still in progress at end of study	2	1	2	1	2	1	2	1	1.5
Targets discontinued	0	1	1	1	1	2	2	2	1.25
TOTAL number of targets	7	7	4	4	6	5	7	8	6

One target accounted for five uses of the discontinue rule, in IDs 2, 4, 5, 7 and 8 (CJ8/9: coordinating noun, verb and adjective phrases using *but not* and *or*). Indeed, no participant reached the 90% correct criterion for this target. The other targets affected only one participant each (ID3: NP1: plural *-s*, ID6: MC9: Subject + Verb + Object + PP and CJ4: coordinated verb and adjective phrases with *and*, ID7: NP9: possessive -*'s* + Noun, ID8: AD4: *if* and *unless*). For ID8 on *if/unless*, we stopped intervention because he was able to produce the target forms reliably in intervention sessions, but in the probe tests, he persisted in asking questions rather than producing a scorable response, despite repeated prompting[4]. Thus, the test did not seem to reflect his ability.

[4] This participant seemed to get into a pattern with the probe test which he persisted with throughout. He asked questions (e.g., "if nobody answers, what shall she do?", "unless the car starts, what shall he do?") instead of producing a statement (e.g., "if nobody answers, she will leave the pizza", "he will call a mechanic, unless his car starts"). For *unless*, the questions were ungrammatical and were therefore scored as zero, but with *if*, the questions were grammatical, but not following the desired structure and therefore un-scorable.

Research Question 5 - Do participants show any changes between the start and end of the study on standardised language tests or other grammatical measures (mean length of utterance, complexity, or errors per utterance)?

We examined whether the participants showed gains on measures of grammar or standardised assessments of grammar and narrative following the study. Note that due to the absence of a control group, we cannot determine whether any progress on these measures is due to intervention or other factors (e.g., maturation, practice effects, general classroom-based language intervention). All participants completed the SPELT-3 (Dawson et al., 2005), the narrative subtest from the ACE (Adams, Cooke, Crutchley, Hesketh, & Reeves, 2001), the 'Beach' Story subtest from the ERRNI (Bishop, 2004) and the TROG-E (Bishop, 2003) both pre- and post-intervention. Summary statistics for all measures are reported in Table 6, while scores for individual participants are shown in Supplementary Information 1, 2 and 3. Our measures of clausal density, mean length of utterance (MLU) and errors per utterance were based on the combined ACE and ERRNI narrative samples for each participant. Clausal density was calculated as the number of clauses in the sample divided by the total number of utterances; MLU was the total number of words divided by the number of utterances, and error rate per utterance was the total number of errors divided by the total number of utterances.

Table 6: Summary statistics for pre- and post-intervention raw scores on grammatical and standardised measures

Measure	Pre-intervention		Post-intervention		Cohen's *d*	Cohen's *d* 95% confidence intervals	*t* value	*p* value
	Mean	SD	Mean	SD				
Clausal density	1.17	0.12	1.27	0.14	0.62	-0.29 – 1.54	1.76	.122
MLU	5.83	1.26	6.26	1.26	0.30	-0.55 – 1.16	0.86	.418
Errors/utterance	0.87	0.77	0.74	0.73	-0.66	-1.59 – 0.26	-1.87	.103
SPELT	28.12	10.47	35.12	6.79	1.07	0.02 – 2.12	3.02	.019
ACE grammar	6.25	2.38	9.12	4.61	0.85	-0.13 – 1.82	2.39	.048
ERRNI MLU	5.51	1.20	6.01	1.26	0.38	-0.49 – 1.24	1.06	.324
TROG	7.12	3.31	8.38	3.58	0.54	-0.35 – 1.43	1.53	.170
ACE information	8.88	4.55	9.75	4.98	0.21	-0.64 – 1.05	0.58	.579
ERRNI content	8.62	2.50	11.00	2.88	0.89	-0.10 – 1.88	2.52	.040

Analysis was conducted in R using the package 'lmerTest' (version 4.2.0), which included raw score as the outcome variable, time point (pre- vs. post-intervention) as the predictor, and by-participant random intercepts. Note that *p* values should be interpreted with caution due to our small sample size: for this reason, we also report standardised effect sizes (Cohen's *d*) and their 95% confidence intervals.

There was no evidence for a statistically significant change in any of the grammatical measures (clausal density, MLU or errors per utterance), although clausal density and errors per utterance had medium effect sizes. For standardised measures of grammar, scores increased significantly on the SPELT and ACE grammar, with large effect sizes, but there was no evidence of change on ERRNI MLU scores or generalisation to receptive grammar on the TROG. We did not expect to observe progress on our standardised measures of

information content, and while this was the case for ACE information scores, there was a large and statistically significant effect for ERRNI content scores.

Discussion

We delivered 1:1 SLT intervention which was highly individualised: specific targets were identified for each participant and short probe tests were used to identify when a target was either achieved or should be discontinued, with the aim of maximising the efficiency of the intervention for each participant and target. In total, between them, the eight participants received intervention on 46 morphosyntactic targets, 27 of which were unique. The results showed that across all participants and targets, scores on probe tests were significantly higher during and after intervention than during the baseline phase. While there was some (marginally significant) progress during the baseline phase, indicating a possible practice effect, the slope during the intervention phase was much steeper and was highly significant,

In contrast, there was no significant change during the maintenance phase, indicating that progress was maintained. The analysis with cumulative teaching episodes as a predictor showed a very similar pattern, with the odds of a correct response increasing by 4.0% for every teaching episode (each of which concluded with a child correct production). Cumulative teaching episodes were shown in our companion paper (REF) to be the best predictor of progress and this measure also has the advantage of including all data from the baseline, intervention and maintenance phases. Thus, in the detailed analyses of the effect of participants, we used cumulative teaching episodes as the main predictor of progress.

When considering the results from both days in intervention phase and cumulative teaching episodes, three of the eight participants in the study showed a different response to intervention in both analyses, with two (ID2 and ID4) showing significantly faster progress with intervention and one (ID6) showing significantly less progress; indeed, for this participant, the intervention does not appear to have been successful. It is of interest that the

two participants who made faster progress were the only two who had been in the school previously and had received similar intervention in the two previous years (focused on different targets), thus they were already familiar with the SHAPE CODING system. It does not seem that progress was related to language level as ID2 had the lowest scores on all three standardised expressive language tests, while ID4 in contrast had either the highest, or second highest score on these tests (see Supplementary Information 1-3). ID6 showed intermediate scores on the standardised tests pre-intervention, but had the lowest attention score (see Table 2).

Three participants showed differing results in the two analyses. ID1 showed steeper progress than the grand mean when using cumulative teaching episodes as the predictor, but not when using days in the intervention phase. ID7 and ID8 showed steeper progress when using days in intervention (but then showed a marginal decline in scores during the maintenance phase) but not cumulative teaching episodes. These different patterns likely result from the fact that cumulative teaching episodes combines results across all three phases. ID1 showed above average performance during both the intervention and maintenance phases. Although this was not significant when the phases were separated, this could have led to significant results once combined. IDs 7 and 8 in contrast both showed declines during the maintenance phase which perhaps cancelled out their stronger progress during the intervention phase when the phases were combined.

In general, the participants appeared to show progress across a range of targets. However, the discontinue criterion due to limited progress was reached for seven of the eight participants for at least one target. One particular target was discontinued in all five participants for whom it was an area of focus, indicating that there was an issue with this target. This target focused on coordinating noun, verb and adjective phrases using *but not* and *or.* Scores on this did show initial progress with intervention, but then performance plateaued. Consideration of the actual items on this probe test showed that most participants made

progress with the items involving coordinated verb and adjective phrases and noun phrases embedded in a verb or prepositional phrase (e.g., *the lady could sing or dance; the car is blue but not fast, the boy wants a cat but not a dog*), but made very limited progress on items involving coordinated subject noun phrases, e.g., *Mum or Dad will wash the dog, the man but not the lady is reading* and *the lady, but not the boy, is hugging the girl*, where they either produced two coordinated clauses ("Mum will wash the dog, or Dad will wash the dog"), or where they added the phrase preceded by *but not* to the end of the clause ("the man is reading, but not the lady", "the lady is hugging the girl, but not the boy" – despite this latter example being ambiguous) . Therefore, in a revised version of the target identification spreadsheet available from www.shapecoding.com, we have now split this structure off into a separate target which has its own separate probe test for use in future studies and clinical practice[5]. In addition, as coordination of the subject with *or* and *but not* is less common in everyday conversations, we recommend this target should not be targeted if participants can understand the structure and are able to express the meaning in another way (e.g., *Mum will wash the dog, or Dad will*, *The man is reading, but the lady isn't*).

The other targets which were discontinued each only affected one participant, and therefore the issue seems to have been more specific to the individual participant and subsets of the target. For example, ID3 and ID7 made good progress with plural *-s* and possessive '*-s* respectively where the added *-s* is realised as [s] or [z], but they made no progress on those items ending in a sibilant, requiring the [ɪz] allomorph, where they continued to make errors, despite specific focus on this during the intervention. We view the use of the discontinue criterion as important for not spending limited intervention time on a target where little (further) progress is being made, especially where there are other structures or areas which could be targeted instead.

[5] We have used the new target codes in this paper. This target therefore has a double two code (CJ8/9) because it has now been split into CJ8 and CJ9.

We also considered whether the participants showed any change on standardised tests, or in more general grammatical measures of length, complexity and accuracy. All measures showed progress in the predicted direction, but it was not possible to judge whether any change on these resulted from the intervention, as we had no control group of other children who were in the same setting but did not receive the intervention. However, due to the nature of the intervention, we predicted more progress on expressive grammatical measures (ACE narrative grammar, ERRNI MLU, SPELT-3) than on measures of information (ACE narrative information, ERRNI content) and receptive grammar, but it is possible that the effects of expressive grammar intervention could extend to these too. Given the small number of participants and hence low power, it was unlikely that we would find significant results. However, we found large and significant changes on the ACE grammar and SPELT-3. In addition, we found medium (but non-significant) effects on clausal density and errors per utterances and an unpredicted large and significant effect on the ERRNI content. This could be due to random variation, activities in the classroom or generalisation.

Limitations and future directions

The small number of participants in this study and the individualised nature of the target identification means that the conclusions from analyses split by participant should be treated with caution, as the results could be affected greatly by the response of an individual participant or by the targets they focused on. The probe test results combining across participants are much more robust as the repeated testing resulted in a large number of data points. However, those analyses which only had two testing points, such as those in research question 5 were underpowered. Thus, further investigation in larger studies is required in order to establish whether the results from the probe tests hold across all targets and a wider range of participants and whether these generalise to standardised tests and general measures of sentence length, complexity and/or accuracy.

Clinical implications

Identifying multiple individualised targets which could all receive similar intervention method has several advantages. It gives clinicians the flexibility to focus on more than one target at a time and also to discontinue a target and start a new one if a participant is not making progress. In addition, participants can use the same intervention method over an extended period of time, and this increased familiarity appears to lead to increased rates of progress.

Conclusions

Our study design enabled us to look in detail at the progress of eight participants on a range of morphological and syntactic targets. In general, the intervention appears to have been highly effective and efficient (with 1-5 targets achieved and another 1-2 still in progress after 8-10 hours of intervention), but progress varied across targets and participants. The results for individual targets have led us to modify the target identification process for the future with the aim of increasing effectiveness.

One participant (with low attention) does not appear to have benefited, but the other seven showed highly significant progress with intervention. The two participants who had received similar intervention previously showed faster progress, indicating that greater gains could perhaps be achieved in a shorter time if the same intervention approach were provided again following this study (but with new targets). We will continue to refine our target identification process and stopping criteria in response to future trials to further improve effectiveness and efficiency and we hope to test the methods further in larger studies in order that generalisation to more general language measures can be better evaluated.

Despite the need for further larger-scale trials, we believe the results of this study suggest that this method is effective for children with DLD who have sufficient attention levels to attend to the SHAPE CODING procedures and feedback hierarchy but who require support with constructing a range of sentence structures accurately. The ability to use the

morphosyntactic forms targeted here will allow children with DLD to express their thoughts and ideas more clearly, will impact their self – efficacy, and is central to both their academic and social functioning.

Acknowledgments

We would like to thank the participants, their families and teachers for their support and flexibility throughout this study.

Data Availability Statement

All data for this study are available from

https://osf.io/32ubp/?view_only=f1ee0fb2a07a4b3198d1b051fa0056b5

References

Calder, S. D., Claessen, M., Ebbels, S., & Leitão, S. (2021). The Efficacy of an Explicit Intervention Approach to Improve Past Tense Marking for Early School-Age Children With Developmental Language Disorder. *Journal of Speech, Language, and Hearing Research, 64*(1), 91-104. doi:doi:10.1044/2020_JSLHR-20-00132

Conti-Ramsden, G., Durkin, K., Toseeb, U., Botting, N., & Pickles, A. (2018). Education and employment outcomes of young adults with a history of developmental language disorder. *International Journal of Language & Communication Disorders, 53*(2), 237-255. doi:10.1111/1460-6984.12338

Curran, M., & Owen van Horne, A. J. (2019). Use of Recast Intervention to Teach Causal Adverbials to Young Children With Developmental Language Disorder Within a Science Curriculum: A Single Case Design Study. *American Journal of Speech-Language Pathology, 28*(2), 430-447. doi:doi:10.1044/2018_AJSLP-17-0164

Ebbels, S. H. (2014). Effectiveness of intervention for grammar in school-aged children with primary language impairments: A review of the evidence. *Child Language Teaching and Therapy, 30*(1), 7-40.

Ebbels, S. H., Maric, N., Murphy, A., & Turner, G. (2014). Improving comprehension in adolescents with severe receptive language impairments: a randomised control trial of intervention for coordinating conjunctions. *International Journal of Language & Communication Disorders, 49*(1), 30-48.

Ebbels, S. H., van der Lely, H. K. J., & Dockrell, J. E. (2007). Intervention for verb argument structure in children with persistent SLI: a randomized control trial. *Journal of Speech Language and Hearing Research, 50*, 1330-1349.

Fey, M. E., Cleave, P., Long, S. H., & Hughes, D. L. (1993). Two Approaches to the Facilitation of Grammar in Children with Language Impairment: An Experimental Evaluation. *Journal of Speech and Hearing Research, 36*, 141-157.

Fey, M. E., Cleave, P. L., & Long, S. H. (1997). Two models of grammar facilitation in children with language impairments: phase 2. *Journal of Speech Language and Hearing Research, 40*, 5-19.

Fey, M. E., Leonard, L. B., Bredin-Oja, S. L., & Deevy, P. (2017). A Clinical Evaluation of the Competing Sources of Input Hypothesis. *Journal of Speech, Language, and Hearing Research, 60*(1), 104-120. doi:10.1044/2016_JSLHR-L-15-0448

Frizelle, P., & Fletcher, P. (2014a). Profiling relative clause constructions in children with specific language impairment. *Clinical Linguistics & Phonetics, 28*(6), 437-449. doi:10.3109/02699206.2014.882991

Frizelle, P., & Fletcher, P. (2014b). Relative clause constructions in children with specific language impairment. *International Journal of Language & Communication Disorders, 49*(2), 255-264. doi:https://doi.org/10.1111/1460-6984.12070

Frizelle, P., Thompson, P. A., McDonald, D., & Bishop, D. V. M. (2018). Growth in syntactic complexity between four years and adulthood: evidence from a narrative task. *Journal of child language*, 1-24. doi:10.1017/S0305000918000144

Frizelle, P., Tolonen, A.-K., Tulip, J., Murphy, C.-A., Saldana, D., & McKean, C. (2021). The Impact of Intervention Dose Form on Oral Language Outcomes for Children With Developmental Language Disorder. *Journal of Speech, Language, and Hearing Research, 64*(8), 3253-3288.

Kulkarni, A. A., Chadd, K. E., Lambert, S. B., Earl, G., Longhurst, L. M., McKean, C., Hulme, C., McGregor, K. K., Cunniff, A., Pagnamenta, E., Joffe, V., Ebbels, S. E., Bangera, S., Wallinger, J., & Norbury, C. F. (2022). Editorial Perspective: Speaking up for developmental language disorder – the top 10 priorities for research. *Journal of Child Psychology and Psychiatry, 63*(8), 957-960. doi:https://doi.org/10.1111/jcpp.13592

Lammertink, I., Boersma, P., Wijnen, F., & Rispens, J. (2017). Statistical Learning in Specific Language Impairment: A Meta-Analysis. *Journal of Speech, Language, and Hearing Research, 60*(12), 3474-3486. doi:doi:10.1044/2017_JSLHR-L-16-0439

Leonard, L. B. (2014). *Children with specific language impairment*: MIT press.

Lum, J., Conti-Ramsden, G., Page, D., & Ullman, M. (2012). Working, declarative and procedural memory in specific language impairment. *Cortex, 48*(9), 1091-1250.

Nippold, M. A. (2010). It's Not Too Late to Help Adolescents Succeed in School. *Language, Speech and Hearing Services in Schools, 41*, 137-138.

Nippold, M. A., Mansfield, T. C., Billow, J. L., & Tomblin, J. (2009). Syntactic Development in Adolescents With a History of Language Impairments: A Follow-Up Investigation. *American Journal of Speech-Language Pathology, 18*(3), 241-251.

Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vamvakas, G., & Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: evidence from a population study. *Journal of Child Psychology and Psychiatry, 57*(11), 1247-1257. doi:doi:10.1111/jcpp.12573

Owen van Horne, A. J. O., Fey, M., & Curran, M. (2017). Do the Hard Things First: A Randomized Controlled Trial Testing the Effects of Exemplar Selection on Generalization Following Therapy for Grammatical Morphology. *Journal of Speech, Language, and Hearing Research, 60*(9), 2569-2588. doi:10.1044/2017_JSLHR-L-17-0001

Plante, E., Mettler, H. M., Tucci, A., & Vance, R. (2019). Maximizing Treatment Efficiency in Developmental Language Disorder: Positive Effects in Half the Time. *American Journal of Speech-Language Pathology, 28*(3), 1233-1247. doi:doi:10.1044/2019_AJSLP-18-0285

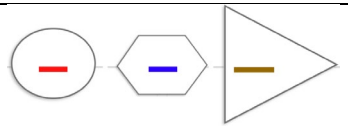
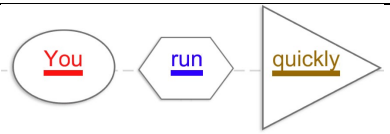
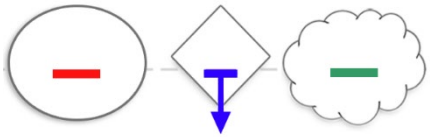
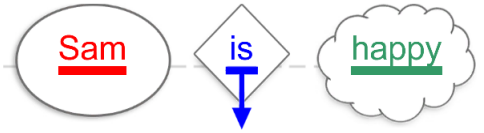
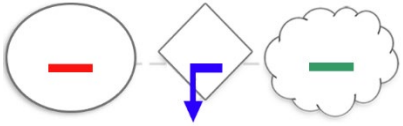
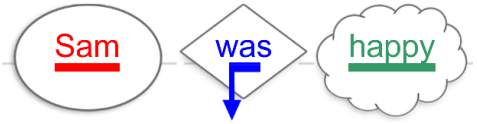
Plante, E., Ogilvie, T., Vance, R., Aguilar, J. M., Dailey, N. S., Meyers, C., Lieser, A. M., & Burton, R. (2014). Variability in the Language Input to Children Enhances Learning in

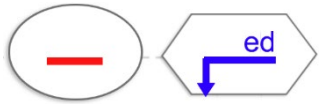
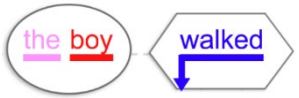
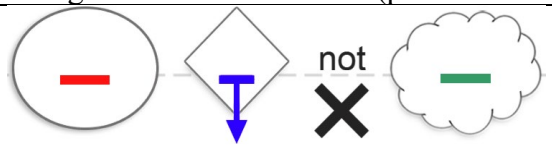
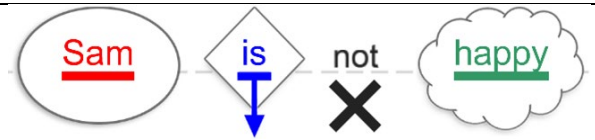
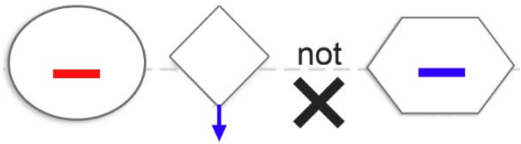
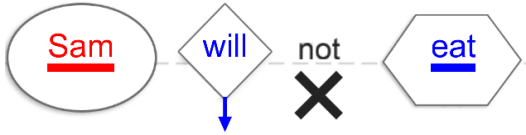
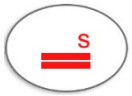
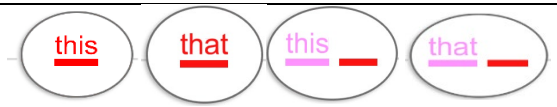
a Treatment Context. *American Journal of Speech-Language Pathology, 23*(4), 530-545. doi:10.1044/2014_AJSLP-13-0038

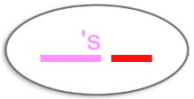
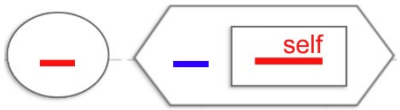
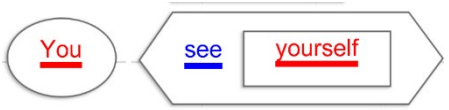
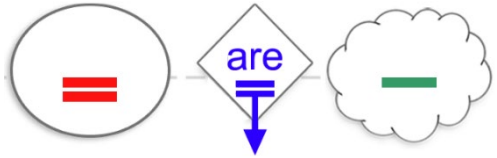
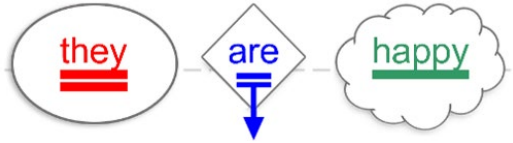
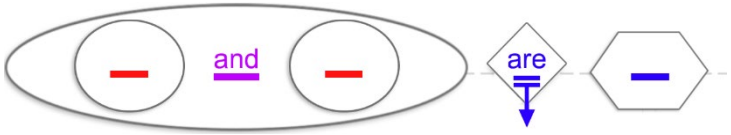
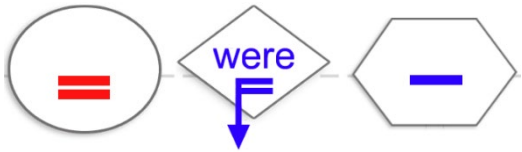
Scott, C. M., & Windsor, J. (2000). General Language Performance Measures in Spoken and Written Narrative and Expository Discourse of School-Age Children With Language Learning Disabilities. *Journal of Speech, Language, and Hearing Research, 43*(2), 324-339. doi:10.1044/jslhr.4302.324

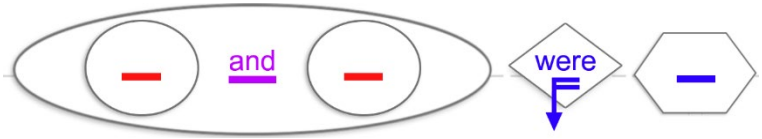
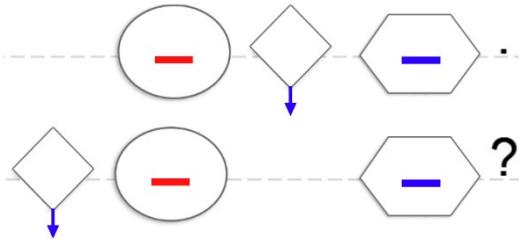
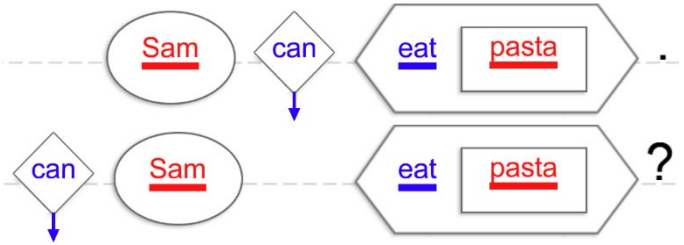
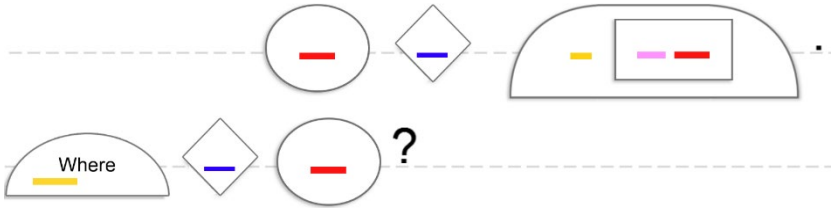
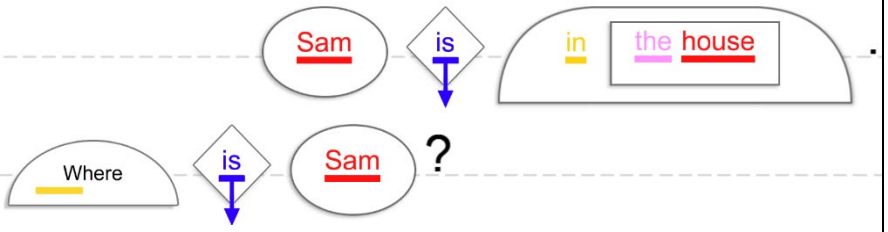
Yew, S. G. K., & O'Kearney, R. (2013). Emotional and behavioural outcomes later in childhood and adolescence for children with specific language impairments: meta-analyses of controlled prospective studies. *Journal of Child Psychology and Psychiatry, 54*(5), 516-524. doi:https://doi.org/10.1111/jcpp.12009

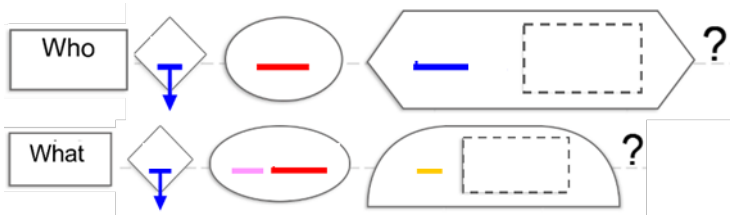
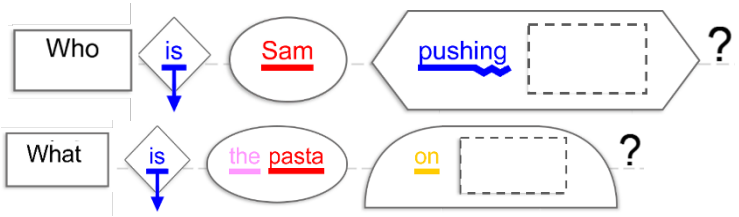
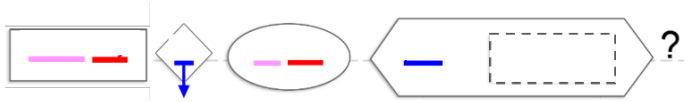
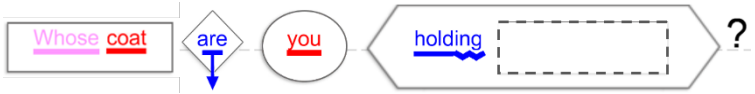
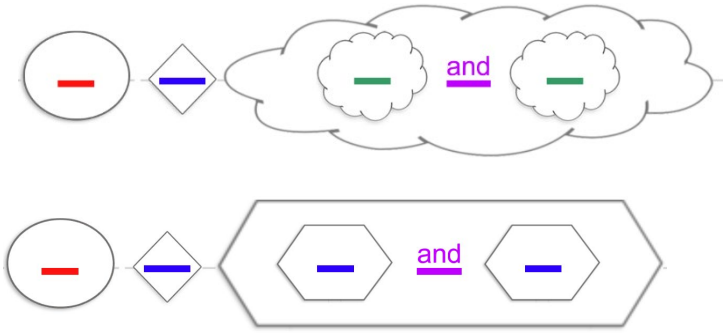
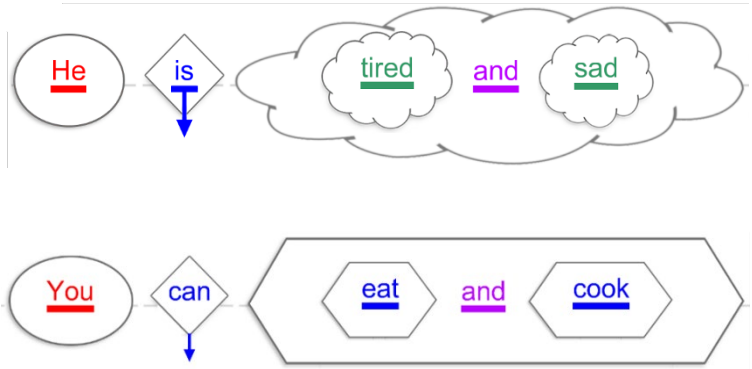
Appendix A – Target codes and descriptions for structures targeted in the study, with SHAPE CODING template, child-friendly rule and a coded example sentence

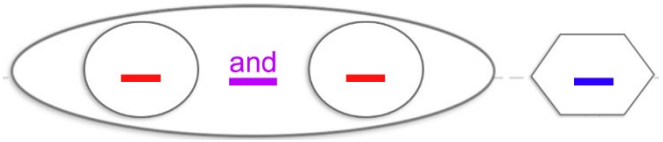
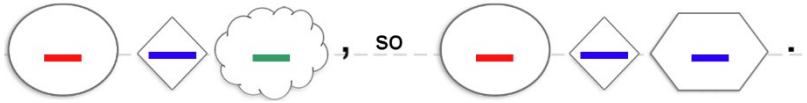
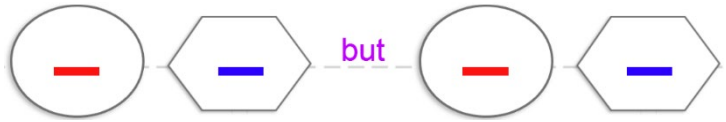
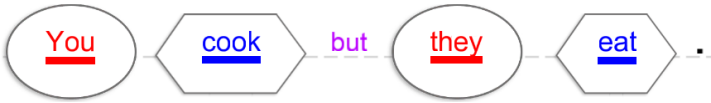
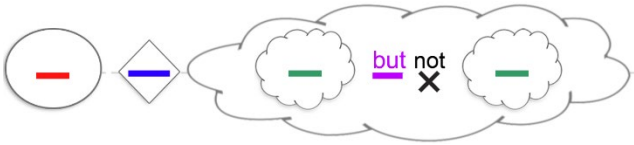
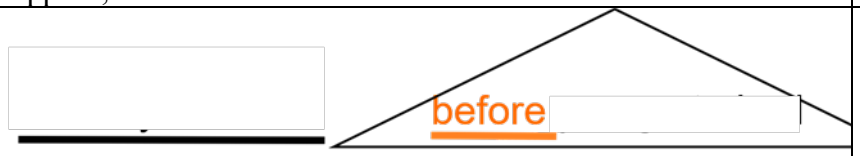
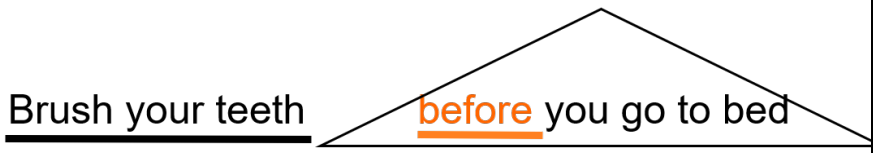
Code: Structure	SHAPE CODING template plus rule	Example
MC9: Subject moves an object to a new place (Subject + Verb + Object + Prepositional Phrase)	Oval moves rectangle to a new place (semi-circle)	
MC10: Adverbs of manner	 Make brown word from green word by adding *-ly*. Brown word tells you how the oval is doing the blue word (pointy triangle goes with pointy hexagon)	
TA2: present tense copula/aux (*is/are/am*)	 Need a blue word (*is, are,* am) in the diamond between oval and cloud	
TA4: past tense copula/aux (*was/were*)		

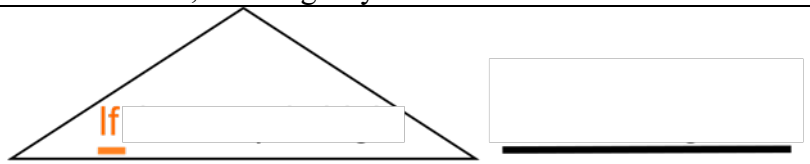
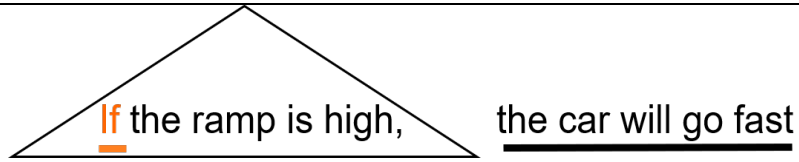
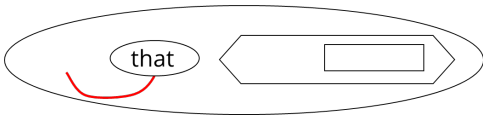
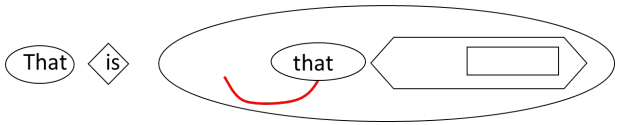
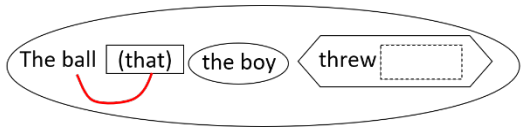
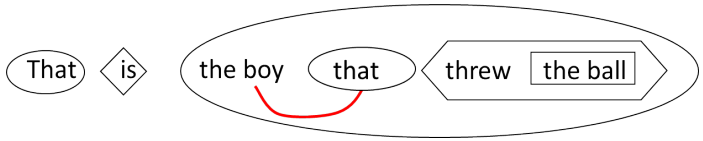
	When talking about past time, we need a past (back) arrow on the blue word in the diamond (this changes *is* and *am* to *was*, and *are* to *were*).	
TA5: sentences requiring the past tense	 Adding back arrow for past time onto hexagon blue word adds *-ed* (pronounced /t, d, Id/)	
NG2: auxiliary/copula + *not*	 The *not* cross goes after a diamond	
NG3: modal + *not*	 The *not* cross goes after a diamond	
NP1: Plural *-s*	 More than one needs two red lines. Add *-s* (pronounced /s,z,iz/)	
NP5: Demonstratives *this* vs *that*	 *This* is for nearby, *that* is for further away. Can be red or pink word	

NP7: Possessive *-s* + Noun	 To show something belongs, add –'*s* to turn red word into pink word	
NP10: Reflexive pronoun	 When oval and rectange are the same person use *myself, yourself, himself, herself, ourselves, themselves* in the rectangle	
AG1: *are* with plural Noun Phrase	 Two red lines in oval needs two blue lines in diamond, *are* in present tense	
AG2: *are* with coordinated Noun Phrases	 Two red lines in big oval (one in each small oval) needs two blue lines in diamond, *are* in present tense	
AG3: *were* with plural Noun Phrase		

	Two red lines in oval needs two blue lines in diamond, w*ere* in past tense	
AG4: *were* with coordinated Noun Phrases	 Two red lines in big oval (one in each small oval) needs two blue lines in diamond, *were* in past tense	
Q11: Question formation with movement of modal	 To ask a *yes/no* question, move the diamond to the front	
Q15: *Where, why, how* questions	 Move the *Wh* shape to the front and then move the diamond to second position	

Q16: *Who, what* object questions requiring movement	 Move the *Wh* rectangle to the front and then move the diamond to second position. To understand these questions, put the rectangle back in place.	
Q19: *Whose, which* Noun object questions requiring movement	 Move the *Wh* rectangle to the front and then move the diamond to second position. To understand these questions, put the rectangle back in place.	
CJ4: Coordinated Verb and Adjective Phrases with *and*	 Join two clouds in a big cloud, or two hexagons in a big hexagon.	

CJ5: Coordinated Noun Phrases with *and*	 Join two ovals together in a big oval	
CJ6: Causal conjunct *so*	 *So* joins two sentences. The first sentence causes the second to happen.	
CJ7: Coordinated clauses with *but, or*	 Join two sentences together with *but*. The second sentence is a surprise.	
CJ8/9: Coordinated phrases with *but not, or*	 Join two shapes the same together with *but not*. The first one happens, the second one doesn't.	
AD3: Adverbial subordinate clauses with temporal conjunctions *before, after, when, until*	 The sentence in the triangle tells you when the main sentence (black line) happens. The main sentence happens 1st with *before*, 2nd with *after*, 2nd with *when* (but straight away, triangle starts it),	

	1st with *until* (triangle stops it). Doesn't matter if triangle appears second or first, meaning stays the same.	
AD4: Adverbial subordinate clauses with conditional conjunctions *if, unless*	 *If* works same as *when* and *unless* same as *until*, difference is they might never happen.	
RC1-4: Unembedded or presentational relative clauses	  Put a whole sentence inside an oval to give more information.	 

774

Supplementary information 1

Individual-level raw and standard scores on standardised measures of grammar

Assessment	ID code	Age at start	Raw		Scaled/Standard[1]	
			Pre	Post	Pre	Post
SPELT	1	9.79	21	32	<40	
	2	9.75	14	31	<40	
	3	8.70	15	25	<40	46
	4	9.61	42	41	89	
	5	8.01	32	33	72	65
	6	9.32	32	32	64	64
	7	10.89	30	42		
	8	8.94	39	45	82	97
ACE grammar	1	9.79	7	4	3	3
	2	9.75	2	2	3	3
	3	8.70	8	14	4	8
	4	9.61	8	10	3	5
	5	8.01	4	7	3	4
	6	9.32	5	9	3	4
	7	10.89	7	15	3	8
	8	8.94	9	12	5	7
TROG	1	9.79	7	7	62	55
	2	9.75	8	10	67	69
	3	8.70	2	6	55	58
	4	9.61	12	14	85	88
	5	8.01	4	4	55	55
	6	9.32	5	6	55	58
	7	10.89	9	13	65	83
	8	8.94	10	7	76	62
ERRNI MLU	1	9.79	5.50	6.17	65	69
	2	9.75	4.10	3.80	64	64
	3	8.70	5.50	6.00	69	69
	4	9.61	6.17	5.80	69	65
	5	8.01	5.47	7.00	69	84
	6	9.32	4.16	5.00	64	64
	7	10.89	5.26	8.04	64	86
	8	8.94	7.92	6.27	90	74

[1]Missing standard scores indicate that norms were unavailable for the child's age at the point of testing. ACE scores are scaled scores (1SD range: 7-13), all others are standard scores (1SD range: 85-115)

Supplementary information 2

Individual-level raw and standard scores on standardised measures of information content

Assessment	ID code	Age at start	Raw		Scaled / Standard[1]	
			Pre	Post	Pre	Post
ACE information	1	9.79	10	6	6	3
	2	9.75	8	4	5	3
	3	8.70	11	12	8	8
	4	9.61	10	9	6	6
	5	8.01	4	8	4	6
	6	9.32	5	6	3	4
	7	10.89	18	19	12	11
	8	8.94	5	14	4	9
ERRNI content	1	9.79	10	10	72	69
	2	9.75	10	11	72	74
	3	8.70	10	11	75	74
	4	9.61	6	12	65	72
	5	8.01	5	11	65	77
	6	9.32	6	5	65	64
	7	10.89	11	15	74	79
	8	8.94	11	13	75	79

[1]ACE scores are scaled scores (1SD range: 7-13), all others are standard scores (1SD range: 85-115)

Supplementary information 3

Individual-level scores on additional measures of grammar

ID code	Measure	Pre	Post
1	Clausal density	1.17	1.27
2		1.08	1.04
3		1.22	1.33
4		1.26	1.32
5		1.12	1.20
6		0.97	1.30
7		1.19	1.53
8		1.36	1.21
1	Error rate utterances	1.00	0.80
2		2.64	2.46
3		0.82	0.72
4		0.53	0.35
5		0.27	0.16
6		0.28	0.59
7		0.92	0.56
8		0.48	0.27
1	MLU	5.93	6.50
2		4.36	3.96
3		5.42	6.26
4		7.42	6.32
5		5.65	6.92
6		4.31	5.56
7		5.73	8.44
8		7.82	6.15

Note. Clausal density is the number of clauses divided by the number of utterances; error rate utterances is the mean number of errors per utterance; MLU is the total number of words divided by the number of utterances. All calculations were made across ACE and ERRNI narratives combined.