Explicit grammar intervention in young school-aged children with Developmental Language Disorder: an efficacy study using single case experimental design

Samuel D Calder[1], Mary Claessen[1], Susan Ebbels[2, 3] and Suze Leitão[1]

Author Affiliations: [1]School of Occupational Therapy, Social Work, and Speech Pathology, Curtin University, Perth, Western Australia, [2] Moor House Research and Training Institute, Moor House School & College, Oxted, United Kingdom, [3]Language and Cognition, University College London, UK

Cite as:

Calder, S., D., Claessen, M., Ebbels, S., & Leitão, S. (2020). Explicit grammar intervention in young school-aged children with Developmental Language Disorder: An efficacy study using single-case experimental design. *Language, Speech, and Hearing Services in Schools, 51*(2), 298-316. doi:10.1044/2019_LSHSS-19-00060

Address correspondence to:

Samuel D Calder

School of Occupational Therapy, Social Work, and Speech Pathology, Curtin University, Perth, Western Australia, GPO Box U1987, 6845.

Email: samuel.calder@postgrad.curtin.edu.au

Conflict of Interest Statement

The authors declare there are no relevant conflicts of interest.

Funding

The authors acknowledge the contribution of an Australian Government Research Training Program Scholarship in supporting this research.

Abstract

Purpose: This study evaluated the efficacy of an explicit combined metalinguistic training and grammar facilitation intervention aimed at improving regular past tense marking for nine children aged 5;10-6;8 years with DLD.

Method: This study used an ABA across participant multiple baseline single case experimental design. Participants were seen 1:1 twice a week for 20-30 minute sessions for 10 weeks and received explicit grammar intervention combining metalinguistic training using the SHAPE CODING™ system with grammar facilitation techniques (a systematic cueing hierarchy). In each session, 50 trials to produce the target form were completed, resulting in a total of 1000 trials over 20 individual therapy sessions. Repeated measures of morphosyntax were collected using probes, including trained past tense verbs, untrained past tense verbs, third person singular verbs as an extension probe, and possessive *'s* as a control probe. Probing contexts included expressive morphosyntax and grammaticality judgement. Outcome measures also included pre-post standard measures of expressive and receptive grammar.

Results: Analyses of repeated measures demonstrated significant improvement in past tense production on trained verbs (8/9 children) and untrained verbs (7/9 children) indicating efficacy of the treatment. These gains were maintained for five weeks. The majority of children made significant improvement on standardised measures of expressive grammar (8/9 children). Only 5/9 children improved on grammaticality judgement or receptive measures.

Conclusion: Results continue to support the efficacy of explicit grammar interventions to improve past tense marking in early school-aged children. Future research should aim to evaluate the efficacy of similar interventions with group comparison studies, and determine whether explicit grammar interventions can improve other aspects of grammatical difficulty for early school-aged children with DLD.

Developmental Language Disorder (DLD) refers to children who experience language difficulties in the absence of known biomedical conditions or acquired brain injury (Bishop, Snowling, Thompson, Greenhalgh, & CATALISE-consortium, 2017). Compared to typically developing peers, children with DLD present with particular difficulties in morphosyntactic skills, such as the use (Rice, Wexler, & Hershberger, 1998) and judgement of grammatical morphemes associated with tense (Rice, Wexler, & Redmond, 1999).

Finiteness marking is challenging for children with DLD (see Leonard, 2014 for a review). Finiteness refers to the obligatory marking of verbs indicating subject-verb agreement and tense, including affixation of morphemes *-ed* (e.g. *the girl walked*) and -S (e.g. *the girl walks*) to verbs for past- and present-tense, respectively. Within English and cross-linguistically, finiteness is a quality of well-constructed clauses (Dale, Rice, Rimfeld, & Hayiou-Thomas, 2018). There is evidence supporting disordered finiteness as a distinct aetiological construct and predictive marker of language growth for DLD (Bishop, Adams, & Norbury, 2006). Children's grammar difficulties are a primary source of parental concern when considering referral for clinical services (Bishop & Hayiou-Thomas, 2008).

Grammar interventions

Treatment for DLD aims to accelerate language growth and remove barriers to functional communication by harnessing strengths (Justice, Logan, Jiang, & Schmitt, 2017). Ebbels's (2014) review indicates an emerging evidence-base for the effectiveness of grammar intervention for school-aged children with DLD. Current evidence is parsed into implicit and explicit approaches to intervention. According to Ebbels's framework, *implicit interventions* target production and understanding of grammar using grammar facilitation techniques implicitly by responding to children's errors in a naturalistic way (Fey, Long, & Finestack, 2003). Children's learning and the knowledge acquired are not necessarily associated with awareness. *Explicit interventions* target increased awareness of the goals of intervention with a pre-established concept of the criteria for success: learning is conscious and deliberate, and information can be recalled on demand (Shanks, Lamberts, & Goldstone, 2005). Within each approach to intervention, specific techniques are used to improve acquisition of grammar.

Implicit interventions using grammar facilitation. Intervention and scaffolding techniques used in implicit approaches are described as *grammar facilitation* (e.g. Fey et al., 2003), which aims to facilitate the acquisition of grammar by increasing the frequency and quality of target forms in input and output. Greater exposure to and opportunities to learn and use language theoretically accelerates the likelihood of language growth (Leonard, 2014). Studies have empirically tested grammar facilitation techniques supporting their use with expressive morphosyntax targets, including imitation (Nelson, Camarata, Welsh, Butkovsky, & Camarata, 1996), modeling (Weismer & Murray-Branch, 1989), focused stimulation (Leonard, Camarata, Brown, & Camarata, 2004), and conversational recasting (see Cleave, Becker, Curran, Van Horne, & Fey, 2015 for a review). Recently, Van Horne, Fey and Curran (2017) reported on a primarily implicit intervention, in which procedures included a combination of sentence imitation, observational modelling, storytelling and focused stimulation, recasting, and cueing for incorrect responses. All 18 four to 10 year old children with DLD enrolled in the study improved their use of regular past tense. Notably, as participants were dismissed from the study following 36 sessions, many still did not achieve mastery of the intervention target. In general, outcomes following implicit intervention are favourable for morphosyntax in preschool-aged children (Leonard, 2014), however, mastery of intervention targets is rarely reported.

Explicit intervention using metalinguistic training. Difficulties with morphosyntax often persist into school age for children with DLD (Bishop, Bright, James, Bishop, & Van der Lely, 2000). An alternative approach may be required because children with DLD may have difficulty learning grammar through implicit grammar facilitation. *Metalinguistic training* aims to improve children's learning of the rules of grammar by creating conscious awareness of grammar through explicit metacognitive teaching (Ebbels, 2014) allowing children to actively reflect on language targets. Meta-awareness is enhanced, so rules of grammar are learned explicitly in a compensatory way.

Metalinguistic techniques can be used explicitly to teach grammar through metacognitive strategies using visual supports and graphic organisers (Ebbels, 2014). The SHAPE CODING™ system is designed to explicitly teach oral and written syntax to children with language disorder (Ebbels, 2007). Ebbels, van der Lely and Dockrell (2007) compared use of the SHAPE CODING™ system with semantic therapy and a no treatment control group with 27 children aged between 10 and 16;1 with DLD. The authors concluded that the SHAPE CODING™ system is a viable and efficacious treatment approach to improve verb-argument structure in older school-aged children. Although evidence for improvement in grammar comprehension is mixed (e.g. Zwitserlood, Wijnen, van Weerdenburg, & Verhoeven, 2015), children may be able to consciously reflect upon the rules of grammar through explicit interventions in the presence of receptive language difficulties to improve understanding, especially older children (Ebbels, Maric, Murphy, & Turner, 2014).

Grammar intervention approaches effective for children above eight years should be tested with younger children to address the concerning gap in evidence for this age group (Ebbels, 2014). Further, Ebbels suggested there may be benefit to integrating therapy techniques to include grammar facilitation and metalinguistic training in a range of activities (e.g. Fey et al., 2003). Combined approaches are yet to be explored extensively.

Combined intervention approaches. In an early-stage efficacy study, Finestack (2018) used a combined implicit/explicit metalinguistic approach compared to an implicit approach to teach novel morphemes to six to eight year old children with DLD. The combined approach was more efficacious than the implicit approach, with gains being maintained and generalised. In a randomised control trial of 31 preschool-aged children, Smith-Lock, Leitão, Prior and Nickels (2015) used explicit teaching principles combined with a systematic cueing hierarchy, which was effective in improving use of expressive morphosyntax when compared to conversational recasting alone. Importantly, the study included a metalinguistic component where children in the explicit group were aware of the therapeutic goal (Smith-Lock et al., 2015). Kulkarni, Pring and Ebbels (2013) conducted a clinical evaluation of the SHAPE CODING™ system combined with elicited production and recasting to improve the use of past tense for two children aged 8;11 and 9;4 with DLD. Both made significant gains in their use of the target structure.

Although grammar facilitation is generally considered implicit (Ebbels, 2014; Fey et al., 2003), there is evidence that the techniques can be used explicitly. In a pilot efficacy study, Calder, Claessen and Leitão (2018) combined the SHAPE CODING™ system with the systematic cueing hierarchy detailed in Smith-Lock et al. (2015) to improve grammar in three children aged seven years with DLD. Importantly, systematic cueing as a grammar facilitation technique in this study was *explicit.* Cues ranged from least to most support, and there was a focus on teaching correct production of grammar through errors to avoid the child perceiving the error to be semantic in nature, as may be the case when using conversational recasting without stating the goal of intervention first. The findings provided early evidence supporting the use of combined intervention approaches to improve receptive and expressive grammar, particularly production of regular past tense following five weeks of intervention. Notably, participants made gains in expressive grammar following only 10 intervention sessions across five weeks, which is markedly shorter duration than reported in many intervention studies. However, the authors acknowledge that including measures of teaching, maintenance and generalisation (e.g. Finestack, 2018) would have broadened understanding of treatment effects, and that a longer period of intervention might be necessary.

Grammar interventions in clinical practice. Recently, Finestack and Satterlund (2018) reported on a national survey of speech language pathology practice in the US. Past-tense verb production was a common intervention goal for practitioners in both early (40%) and elementary education settings (60%). Interestingly, overall between 60-70% used explicit presentations as an intervention procedure, despite relatively little investigation in this area until recently. Therefore, it appears explicit instruction to improve past tense may not only be supported by an emerging evidence-base, but is also frequently used in clinical practice.

The current study

For early school-aged children, preliminary data suggest that explicit combined metalinguistic and grammar facilitation approaches are efficacious in treating the use of tense marking and for improving receptive grammar more generally (Calder et al., 2018). Building on early stage studies of treatment efficacy is required to determine if treatment procedures are considered evidence-based. Fey and Finestack (2008) outline the need for a programmatic approach to pursuing intervention research, specifically noting the value of small scale studies aimed at exploring and identifying specific components of intervention approaches and their effects on specific populations. This study forms a part of a programme of research to design, develop and evaluate the efficacy of an explicit combined grammar intervention in line with Robey's Phases of Clinical Research (Robey, 2004). We report on a range of measures to evaluate the efficacy of explicit intervention to improve grammar. Single case experimental design (SCED) methodology was used to test the following confirmatory hypotheses and is reported as per the Single-Case Reporting Guideline in Behavioural Interventions (SCRIBE) (Tate et al., 2016):

1. For young school-aged children (specifically, 5;10-6;8 years) with DLD, there will be a significant treatment effect on trained past tense verbs, and a generalised effect to

untrained verbs across 20 sessions of explicit intervention combining metalinguistic and grammar facilitation techniques.

2. These children will improve significantly on pre-post standardised measures of expressive and receptive grammar.

Method

Research Design

Design. The current study was an ABA across participant multiple baseline single case experimental design (SCED) including a minimum of five data points (i.e. sessions) for each phase (Kratochwill et al., 2012). Multiple baselines were conducted for varied durations across participants, and introduction of treatment to participants was staggered. Repeated measures were collected throughout the intervention phase and post-treatment maintenance phase (Dallery & Raiff, 2014), including the target behaviour (past tense verbs), an extension of the targeted behaviour (third person singular verbs) and a control behaviour (possessive *'s*). This design is noted for robustness regarding strengths of internal validity and external validity when compared to other SCEDs (Tate, et al., 2016). As a Phase I-II study, we replicated and built on findings from Calder et al. (2018) by refining intervention protocols, determining optimal dosage and evaluating duration of therapeutic effect (Robey, 2004).

Randomisation. To improve internal validity further, participants were randomly assigned to one of three pre-determined staggered onset to intervention conditions. To ensure concealed allocation, participants were assigned a code which was entered into a random list generator by a blinded researcher. Participants received: five (P1, P3, P8), seven (P5, P7, P9) or nine (P2, P4, P9) pre-intervention baseline sessions over as many weeks; 20 intervention sessions over 10 weeks, and; five post-intervention sessions to evaluate maintenance. Participants were also randomised to grammaticality conditions described below.

Blinding. Participant caregivers and teachers were aware children were receiving grammar intervention but were blinded to the intervention target. Post-intervention measures were collected via blinded assessment using trained student speech-language pathologists.

Participants

Selection criteria. Participants included nine early school-aged children diagnosed with DLD. The inclusion criteria were: aged between 5;6 and 7;6; English as a primary language, and; grammar difficulties associated with DLD. Exclusionary criteria included: a neurological diagnosis, a cognitive impairment, and hearing outside normal limits. Participants were recruited from a specialised educational program for students diagnosed with DLD. Ethical approval for the study was obtained from the Curtin University Human Research Ethics Committee (Approval number: HRE2017-0835) and the Western Australian Department of Education. The principal consented school participation and then provided information letters and consent forms to the parents/carers of potential participants identified by speech-language pathologists and teachers employed at the educational program. Parents returned the completed consent forms if they wished their child to participate. The study reached capacity at nine participants so we could achieve three replications over three baseline conditions as per reporting standards (Kratochwill et al., 2012).

Participant characteristics. The participants' school enrolment package was accessed, including the assessment protocol and the most recent standardised assessment scores available. Data included Clinical Evaluation of Language Fundamentals Preschool- 2 (Wiig, Secord, & Semel, 2004); a test of non-verbal IQ, and; a comprehensive exploration of previous medical history to identify contributing factors to language difficulties, such as

Table 1

Demographic information

Participant ID	Sex	Age at enrolment to school (year; month)	Current year at specialised educational program	Age at initial assessment for study (year; month)
P1	Male	4;0	3rd	6;3
P2	Male	3;11	3rd	6;2
P3	Male	4;7	2nd	5;10
P4	Male	5;4	3rd	6;8
P5	Male	5;2	2nd	6;6
P6	Female	5;11	1st	6;2
P7	Male	5;3	2nd	6;7
P8	Male	3;8	3rd	6;0
P9	Male	4;9	2nd	6;1

acquired neurological damage, or hearing loss. These factors combined are considered evidence of a diagnosis for DLD (Bishop, Snowling, Thompson, Greenhalgh, & CATALISE-consortium, 2016). Participants then passed a hearing acuity test. All participants passed the Phonological Probe from the Test of Early Grammatical Impairment (Rice & Wexler, 2001) for articulation of phonemes necessary for morphosyntactic production targets.

All demographic information is presented in Table 1. Participants included eight males and one female aged between 5;10 and 6;8 at initial assessment. Ages at enrolment to the specialist school varied from 3;8 years to 5;11. P1, P2, P4 and P8 were in their third year of placement at the school; P3, P5, P7 and P9 were in their second, and; P6 was in her first.

Measures

Repeated Measures.

Repeated measures of morphosyntax were collected at every data point using various probes, including: trained probes, untrained probes, an extension probe and a control probe (elaborated in the following sections). Probing contexts included both expressive morphosyntax and grammaticality judgement. Grammaticality judgement was selected as a method of measuring grammatical progress, as there is evidence performance on such tasks mirrors production tasks (Rice et al., 1998; 1999). As grammaticality judgement is a clinical marker of DLD (Rice et al., 1999; Dale et al., 2018), identification of grammatically correct sentences in the studied participants was expected to be below chance levels of accuracy prior to intervention.

Trained probes. Regular past tense (*-ed*) repeated measures of trained verbs were probed in two conditions: 12 *-ed* verbs trained within sessions were measured, and; 12 *-ed* verbs from the previous session were measured. All *-ed* verbs were predetermined at the outset of intervention and selected based on their suitability to intervention activities. We also chose verbs that were not in the Grammar Elicitation Test (GET; described below; Smith-Lock, Leitão, Lambert, & Nickels, 2013) to allow comparison between trained and untrained verbs. These probes were administered during the intervention phase at the end of session 2 (i.e. data point B1 the first week of intervention), and every even session thereafter.

Untrained probes. Repeated measures of untrained expressive morphosyntax probes were selected from an adapted version of the GET. This experimental test was designed to elicit multiple instances of specific expressive morphosyntax targets, including 30 items probing the treated grammatical structure (*-ed*). Repeated measures were also developed for a grammaticality judgement task including 30 *-ed* probes. Videos of actions depicting the declarative clauses containing *-ed* were created as stimuli for untrained probes. Accompanying audio for each task item, both grammatical and ungrammatical (e.g. *The girl painted a picture.* vs. *The girl paint* a picture.*) was recorded by an adult female with an Australian accent, blinded to the purpose of research. Each video with corresponding audio was embedded into a Microsoft PowerPoint presentation. Participants wore Sony noise-cancelling headphones during administration and were required to decide if the sentence 'sounded right' by pressing 'yes' or 'no' on a tablet app. Items were counterbalanced for grammaticality so participants did not receive the same combination of grammatical/ungrammatical items, and there was no pattern in presentation of grammatical/ungrammatical items to counteract a priming effect.

Complete sets of 30 untrained *-ed* verbs were probed pre- and post-intervention. Sets were randomised for administration at the initial assessment (Timepoint 1), one week prior to intervention commencing (Timepoint 2), one week following intervention (Timepoint 3) and five weeks following cessation of intervention (Timepoint 4). Both expression and grammaticality judgement were assessed.

Reduced randomised sets were generated for each other data point using nine expressive probes and 12 grammaticality judgement probes. All possible allomorphs were included (i.e. [d], [t] and [əd]) and equally distributed. Probes were administered via laptop during the pre-intervention baseline phase, at the beginning of session 3 (i.e. data point B2 in the second week of intervention) and every odd session thereafter during the intervention phase, and in the post-intervention maintenance phase.

Extension probes. Expressive repeated measures of third person singular (3S) served as an extension of the treated structure. Items included 30 probes and were taken from the GET. A grammaticality judgement task was also developed as per the untrained *-ed* probes (e.g. *The man sneezes*. vs. *The man sneeze*.*). 3S was considered an extension measure due to the structure's relative complexity compared to *-ed*, since bare stem forms are grammatical when used with first person subject pronouns or plural subject nouns (e.g. *I like ice-cream* vs. *The boys like ice-cream* vs. *The boy likes ice-cream*). We also expected there might be improvement in 3S due to the frequent instances of input during therapy (see Intervention section) and increased awareness of the need for tense marking.

Control probes. Similarly, expressive repeated measures of possessive *'s* (*'s*) served as a control probe. Items included 30 probes and were taken from the GET. As above, a grammaticality judgement task was developed (e.g. *The spider is living on a leaf. This is the spider's leaf.* vs. *The spider is living on a leaf. This is the spider* leaf.*). For *'s,* still images of nouns depicting ownership were retrieved from copyright free image sources. *'s* was considered a control as this noun possession was not taught as part of therapy and therefore should remain stable throughout the intervention period.

For extension and control probes, all possible allomorphs were included (i.e. [s], [z], [əz]) and equally distributed. Randomised sets of 9 expressive and 12 grammaticality judgement items were generated and administered as per the untrained *-ed* probes during pre-intervention, intervention, and post-intervention phases.

Pre-post.

The Structured Photographic Expressive Language Test 3rd Edition (SPELT-3) (Dawson, Stout, & Eyer, 2003) and the Test of Reception of Grammar 2nd Edition (TROG-2) (Bishop, 2003) were administered both pre- and post-intervention as expressive and receptive standardised grammar measures, respectively. The SPELT-3 measures expressive morphosyntax using 54 items across a range of structures and was normed on children aged four to nine years. To address discriminant accuracy of the test, Perona, Plante and Vance (2005) determined 90% sensitivity and 100% sensitivity at 95 cutoff (-0.33SD). This cutoff score was used for the current study based on the recommendation, although it is noted that while other studies applied this cutoff with older children (e.g. Van Horne et al., 2017), Perona et al. (2005) sampled children aged four to five years. The TROG-2 test measures a total of 20 different grammatical structure contrasts and was normed on children aged four to 16. Discriminant accuracy was evaluated on a sample of 30 children aged 6;2-10;11 which confirmed the test is sensitive to identifying communication difficulties in children (Bishop, 2003). Both tests have strong reliability and appropriate validity.

Reliability.

A blinded researcher scored 20% of all measures audio and video recorded throughout the study. Inter-rater reliability for experimental measures was calculated using intraclass correlation coefficients (ICC) using absolute agreement and single measures in a two-way mixed effects model. Interpretation of ICC values are as follows: <.40 = poor; .40-.59 = fair; .60-.74 = good, and; .75-1.00 = excellent (Cicchetti, 1994). For trained *-ed* probes, the ICC for expressive measures was .879 and ICC for grammaticality judgement was .977. ICC for expressive untrained -*ed*, 3S and *'s* probes was .937, and ICC for the grammaticality judgement of untrained -*ed*, 3S and *'s* was .985. Therefore, excellent agreement was observed across all experimental measures.

Intervention

All intervention sessions were videotaped and carried out in a quiet space at the site of the educational program. Procedures were similar to those reported by Calder et al. (2018) and are explained within the model suggested by Warren, Fey, and Yoder (2007) for describing treatment intensity. The dose was 50 trials within 20-30 minute sessions; dose form was explicit intervention combining metalinguistic training using the SHAPE CODING™ system (Ebbels, 2007) with a systematic cueing hierarchy (Smith-Lock et al., 2015); dose frequency was twice a week; total intervention duration was 10 weeks, and; cumulative intervention intensity was (50 trials x 2 times per week x 10 weeks), resulting in a total of 1000 trials over 20 individual therapy sessions through roughly 7-10 hours of therapy. This is double the intervention duration in the pilot study (Calder et al., 2018), where authors suggested that participants may demonstrate larger treatment effects following a longer duration. Training of morphosyntax was embedded within engaging and naturalistic activities suited to early school-aged children, including playdough, board games, and playing with puppets, and farm and sea creature manipulatives. Target morphemes were presented in syntactic structures as they occurred felicitously within these activities. The first author (SDC), a trained speech-language pathologist (SLP), delivered all intervention.

Each session began with a short recap of the aims: to say WHAT DOING words (verbs) that have already happened, and to add the sounds ([d], [t], [əd]) onto the end of those words. Next, the SLP would direct the child's attention to the laminated shapes and arrows used as a visual organiser throughout session activities. See Figure 1 for essential shapes,

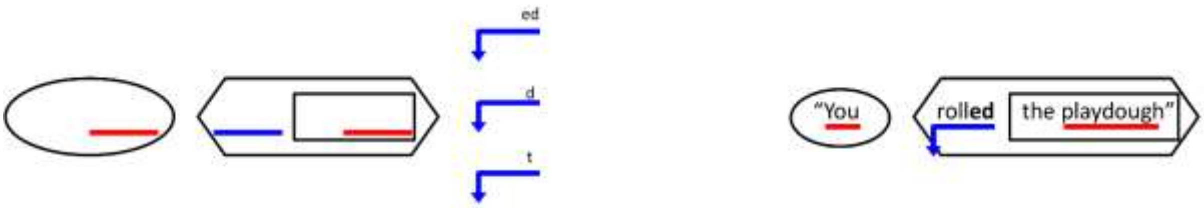

Figure 1. Visual depiction of visual cues used during intervention phase.

including the oval (subject noun phrase WHO/WHAT?), the hexagon (verb phrase WHAT DOING?) and the rectangle (object noun phrase WHO/WHAT?). Additional visual cues included three separate laminated cards that depicted a ‘left down arrow’ to depict *-ed*, and an orthographic representation of the allomorphs (i.e. ‘d’ for [d], ‘t’ for [t], and ‘ed’ for [əd]). The SLP said, *“Last time, we used our shapes and arrows to help us. Like this: ‘We move our shapes and arrows. What did we do? We move***d** [bring ‘d’ arrow into the WHAT DOING? hexagon] *our shapes and arrows. The [d] at the end of move***d** *lets us know it’s already happened.”* The participant was reminded, *“I (the SLP) will say what we do in the session (i.e. present tense) and you will say what we did (i.e. past tense)”.* This was followed by two activities which were designed to give the participants ample opportunities to produce *-ed* verbs in response to an interrogative (e.g. *What did you do?; Did you just VERB? Tell me*…).

Each activity began with explicit instruction of how to apply *-ed* inflection, using one exemplar from each of the allomorphic categories. Within each activity, there were approximately 25 opportunities for the child to respond to an interrogative (e.g. *You roll the playdough! What did you do?)* using *-ed* verbs while the SLP gestured to the shapes and arrows (see Figure 1). The child was therefore encouraged to respond using a Subject-Verb-Object syntactic frame, consistently. If the child responded with an unmarked verb (i.e. bare stem) or overgeneralised form (e.g. *playded*), s/he was supported with a systematic cueing hierarchy moving from least to most support outlined in Figure 2. As much as possible, verbs were blocked according to allomorphs and presented from least to most difficult (i.e. [d]→[t]→[əd]) in accordance with Leonard (2014) and Marshall and van der Lely (2006). At the end of every activity, the SLP recapped what the participant had learned using the shapes and arrows, and comprehension questions. For example, if the target sentence had been '*I rolled playdough'*, the SLP would gesture to the WHO?/WHAT? oval and ask, *"WHO rolled the playdough?"* Then gesture to the WHAT DOING? hexagon while bringing down the 'd' left down arrow and ask, *"What DID you DO?"*, and finally gesture to the WHO?/WHAT? rectangle and ask, *"WHAT did you roll?"* Plausible responses to all of these questions are '*I rolled the playdough'*, giving further opportunity to reinforce production using a consistent syntactic frame. If an error occurred, the same systematic cueing hierarchy described above was employed. The shapes and arrows were then removed, and the interrogative (*What DID you DO?*) was repeated without visual support for an exemplar from all three allomorphic categories, reinforcing internalisation of the grammatical rule. If a child had achieved 80% success over three sessions on any measure, *'s*illy Sentences' were introduced; a metalinguistic sub-activity whereby three sentences where said, either grammatically or ungrammatically (i.e. -*ed* morphemes were either included or omitted), and the child would decide if the sentence 'sounded right'.

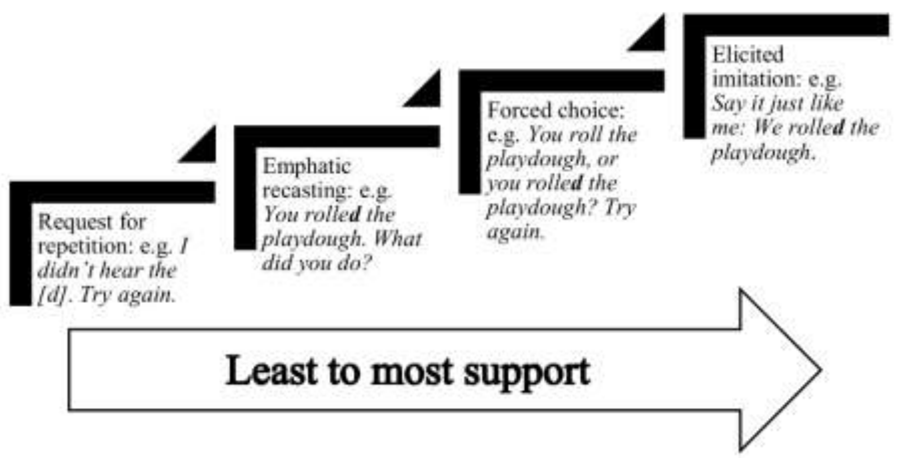

Figure 2. Systematic cueing hierarchy used when child produced the target verb in error.

These procedures were repeated for a second activity, giving 50 opportunities to use -*ed* inflection during the activity which was bookended with explicit teaching and comprehension questions using three exemplars from each allomorphic category. At the end of each session, the child was reminded of the goal of the session, and why it is important to say the sounds at the end of '*WHAT DOING?'* words that have already happened, and also to listen out for those sounds.

Procedural fidelity.

A blinded researcher scored 20% of videotaped sessions on percentage accuracy using *a priori* established criteria for intervention procedures. A total of 19 items were scored for sessions (see Appendix A for a checklist for scoring intervention procedure fidelity). Note, if children were introduced to *'s*illy Sentences', sessions were scored against an additional two (total 21) items. Intra-observer agreement was calculated using ICC. The average score was 97.1% for percentage accuracy, and ICC for treatment procedures was .996.

Analysis

Single subject analyses. Treatment effects of teaching, generalisation and maintenance through repeated measures of morphosyntax were statistically evaluated using *Tau-U* by combining non-overlap and trend of data (Parker, Vannest, Davis, & Sauber, 2011) across all phases and data points. *Tau-U* uses Kendall's *S* to interpret significance testing and outputs *p* values. Raw scores on probes were converted to percentage correct. Baselines were contrasted using the *Tau-U* online calculator (Vannest, Parker, Gonen, & Adiguzel, 2016), and the *Tau* value was checked for trend of baseline in pre-intervention and post-intervention phases. For pre-intervention baseline, *Tau* values above 0.40 (increasing trend) or below -0.40 (decreasing trend) were deemed unstable and corrected, as recommended by Parker et al. (2011). This was repeated for all applicable baseline versus intervention contrasts. Finally, phase contrasts were aggregated to provide an omnibus effect size for study participants, where, using Cohen's standard, 0.2 is small, 0.5 is medium and 0.8 is large.

To evaluate performance on the full sets of untrained *-ed* verbs, a concurrent within-group approach was used (e.g. Zwitserlood et al., 2015) where Friedman non-parametric two-way analysis of variance (ANOVA) tested differences between Timepoint 1 and 2 pre-intervention, and Timepoint 3 and 4 post-intervention scores. Participant scores determined a group mean and standard deviation in expressive and grammaticality judgement probes within each Timepoint. Post-hoc Wilcoxon sign-rank tests made pairwise comparisons between testing points. These statistics were computed using IBM SPSS Version 25.

Kratochwill et al. (2012) outline standards for analysis of repeated measures via visual inspection to report on a functional relation between dependent and independent variables, which includes comments on level, trend and variability within phases, and comments on immediacy, overlap and consistency between phases. For the current study, within phase level performance was evaluated with group statistics. Further, *Tau-U* handles within phase level, and trend and variability within *and* between phases, as well as overlap between phases. Therefore, reporting on visual inspection is limited to the immediacy of the functional relation between *-ed* use and understanding, and the staggered introduction of intervention across participants.

Pre-post analyses. Pre-post differences on standardised measures were tested in a case series approach by calculating the Reliable Change Index (RCI) (Unicomb, Colyvas, Harrison, & Hewat, 2015). The RCI statistic calculates whether an individual's change in score (i.e. pre-post difference in standard scores) is statistically significant by using the reliability values of a standardised test. The RCI is calculated using the formula x_2-x_1/S_{diff}, where x_1 is the participant's pre-test score, x_2 is the same participant's post-test score, and S_{diff}

is the standard error of difference between the two test scores. An RCI above 1.96 is considered statistically significant at 0.05 significance level.

Results

Sequence completed

All participants completed planned sessions within pre-intervention baseline (A), intervention (B), and post-intervention maintenance (A) phases. There was an average of 50.74 (SD= 1.2; range 48-56) trials for each participant to produce *-ed*. Out of the nine participants, six (P1, P2, P3, P4, P5, P7) demonstrated at or above 80% performance on at least one measure of *-ed* marking over three sessions. These participants engaged in the 'Silly Sentences' aspect of intervention procedures as described in the Intervention section.

Outcomes and estimation

Single subject treatment effects (expressive). Data not reported in tables are available in Supplementary Materials (https://asha.figshare.com/articles/Grammar_intervention_in_young_children_with_DLD_Calder_et_al_2020_/11958771). Pre-intervention baselines on production of *-ed* verbs taken from the GET were stable for 4/9 participants. P1 (*Tau* = -0.70), P3 (*Tau* = -0.70), P4 (*Tau* = 0.58), P8 (*Tau* = 0.60) and P9 (*Tau* = -0.71) had baselines corrected for subsequent analyses. Data from expressive repeated measures are presented in Figures 3-5 and results from *Tau-U* analyses are reported in Table 2. Of the nine participants, eight (P1-P7, P9) demonstrated statistically significant trend in production of trained verbs tested within-session during the intervention phase (Figure 3). Phase contrasts were combined and yielded an aggregated effect size of 0.88, which is considered large. For trained verbs tested between sessions (Figure 4), seven (P1-P5, P7, P9) of the nine participants demonstrated statistically significant performance during the intervention phase with a large aggregated effect size of

0.83. Seven (P1-P7) of the nine participants demonstrated a statistically significant trend in production of untrained *-ed* verbs during the intervention phase (Figure 5) yielding a medium effect size of 0.64.

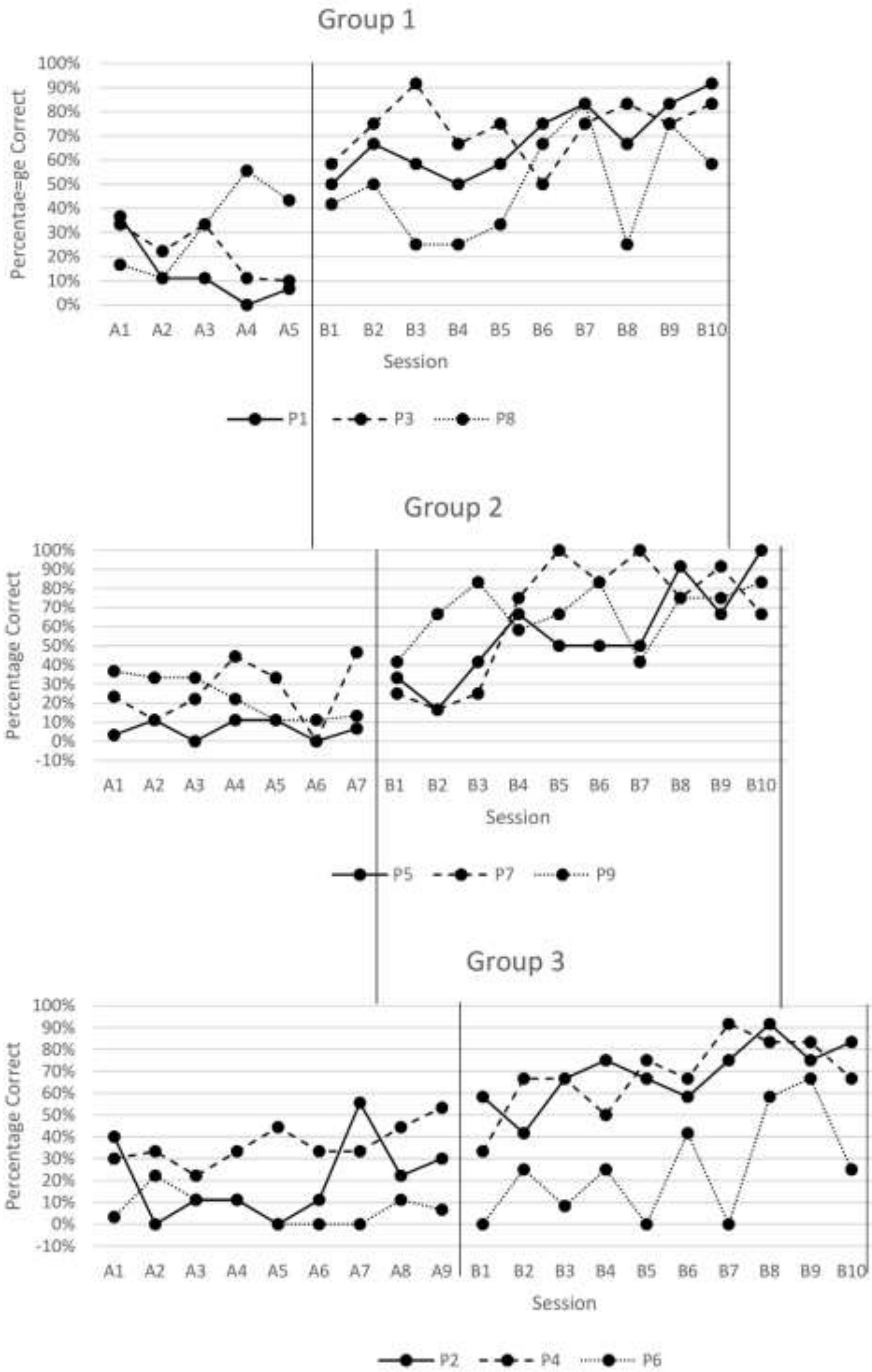

Figure 3. Percentage correct on expressive trained within-session probe repeated measures for Groups 1-3.

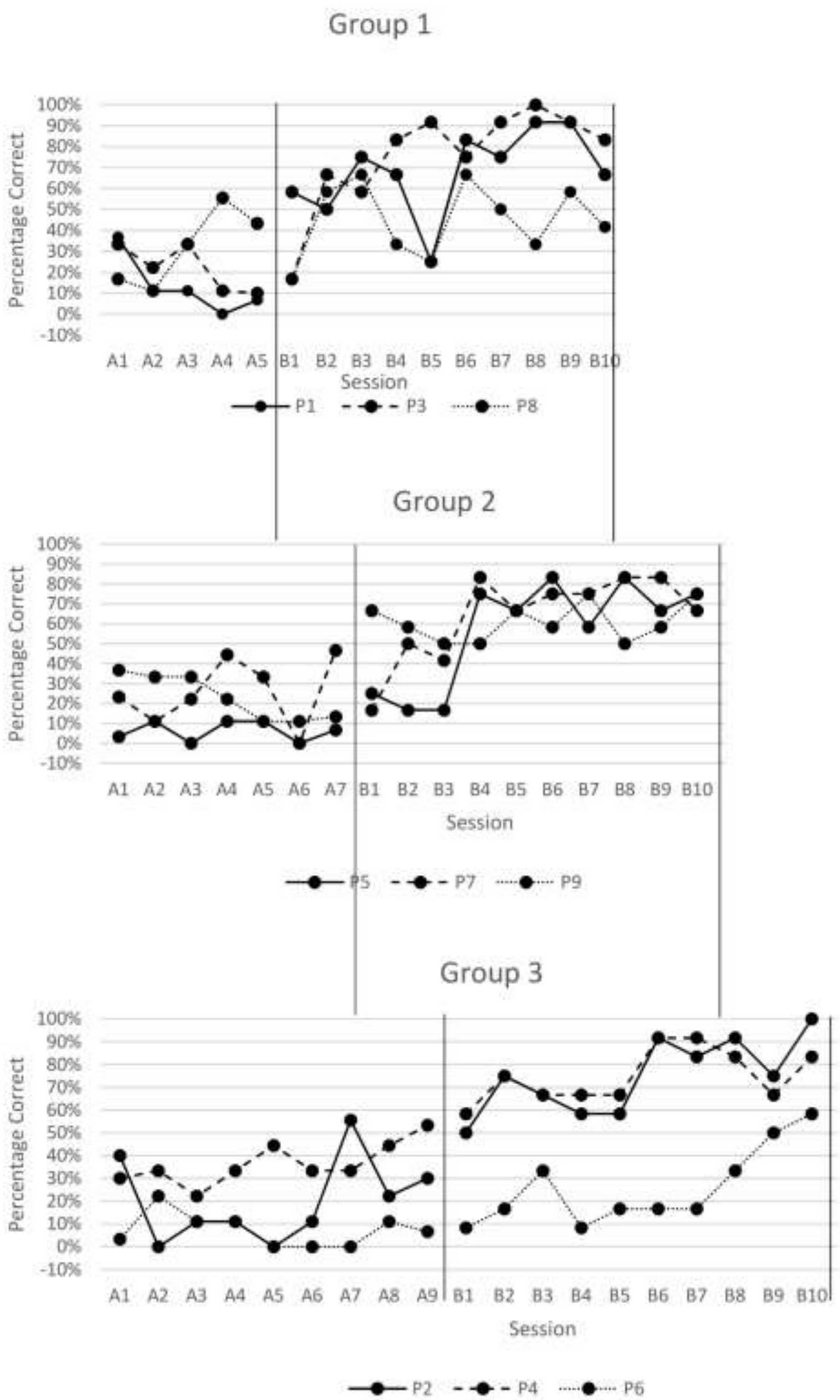

Figure 4. Percentage correct on expressive trained between-session probe repeated measures for Groups 1-3.

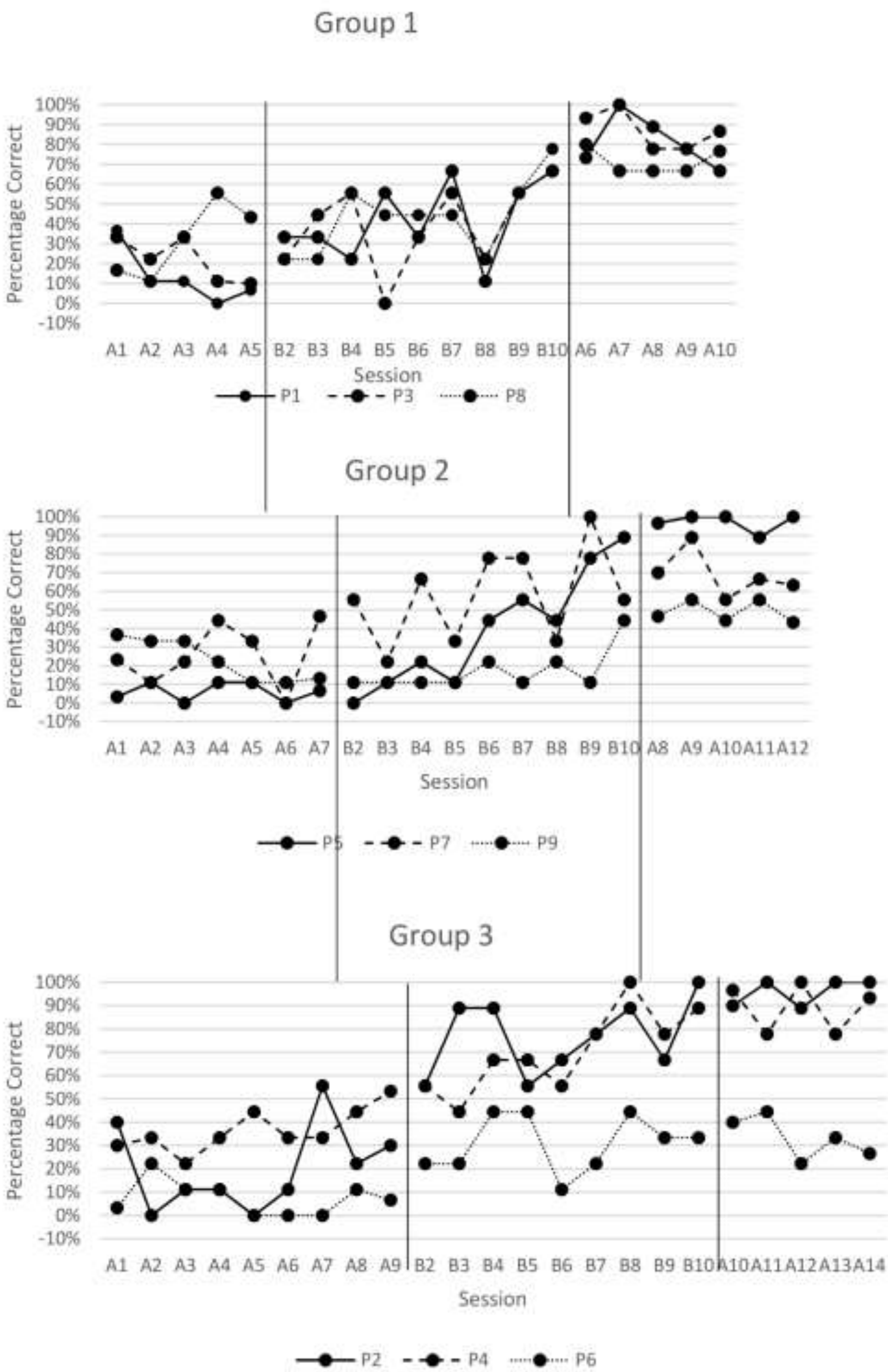

Figure 5. Percentage correct on expressive untrained probe repeated measures for Groups 1-3.

Table 2

Summary of expressive repeated measures baseline versus treatment phase contrasts on trained and untrained targets

Participant ID	Kendall's *S*	*z* score	*p* value	*Tau*	90% CI
WITHIN SESSION					
P1[a]	55	3.37	<0.001*	1.1	[0.56, 1]
P2	88	3.60	<0.001*	0.98	[0.53, 1]
P3[a]	51	3.12	0.002*	1.02	[0.48, 1]
P4[a]	69	2.82	0.005*	0.77	[0.32, 1]
P5	70	3.42	<0.001*	1	[0.52, 1]
P6	66	2.70	0.007*	0.73	[0.29, 1]
P7	56	2.73	0.006*	0.80	[0.32, 1]
P8[a]	15	0.92	0.36	0.30	[0.24, 0.84]
P9[a]	85	4.15	<0.001*	1.21	[0.73, 1]
				Aggregated ES	
Group	-	-	<0.001*	0.88	-
BETWEEN SESSION					
P1[a]	57	3.49	<0.001*	1.14	[0.60, 1]
P2	88	3.59	<0.001*	0.98	[0.53, 1]
P3[a]	57	3.49	<0.001*	1.14	[0.60, 1]
P4[a]	57	2.33	0.02*	0.63	[0.19, 1]
P5	70	3.42	<0.001*	1.00	[0.52, 1]
P6	37	1.51	0.13	0.41	[-0.04, 0.86]
P7	48	2.34	0.02*	0.69	[0.20, 1]
P8[a]	15	0.92	0.36	0.30	[-0.24, 0.84]
P9[a]	85	4.13	<0.001*	1.21	[0.73, 1]
				Aggregated ES	
Group	-	-	<0.001*	0.83	-
UNTRAINED					
P1[a]	40	2.67	0.007*	0.89	[0.34,1]
P2	79	3.49	<0.001*	0.98	[0.52, 1]
P3[a]	30	2.00	0.05*	0.67	[0.12, 1]

P4[a]	56	2.47	0.01*	0.69	[0.23, 1]
P5	45	2.38	0.02*	0.71	[0.22, 1]
P6	73	3.22	0.001*	0.90	[0.44,1]
P7	44	2.33	0.02*	0.70	[0.21, 1]
P8[a]	13	0.87	0.39	0.29	[0.26, 0.84]
P9[a]	-8	-0.42	0.67	-0.13	[0.62, 0.37]
				Aggregated ES	
Group	-	-	<0.001	0.64	-

Notes. CI= confidence interval; ES= effect size
*sig.
[a]unstable baseline corrected

Analysis of *Tau* scores revealed a significant negative trend in performance for P1 (*Tau* = -0.40), P6 (*Tau* = -0.40) and P7 (*Tau* = -0.40) across five datapoints in the post-intervention maintenance phase. Note the *Tau* values for these three participants is at minimum level for baseline trend (*Tau* = ±0.40) corrections according to Parker et al. (2011).

For expressive 3S extension probes, P7 (*Tau* = 0.62), P8 (*Tau* = 0.60) and P9 (*Tau* = 0.57) demonstrated an unstable baseline with a positive trend. During the intervention phase, P6 demonstrated significant improvement (p = .03) and P9 demonstrated significant decline (p = .03). Phase contrasts yielded a non-significant (p = .65) aggregated effect size of -0.05. P1 (*Tau* = 0.80), P2 (*Tau* = 0.40) and P4 (*Tau* = 0.70) demonstrated positive trend in the post-intervention maintenance phase.

For expressive *‘s* control probes, P2 (*Tau* = 0.69) and P4 (*Tau* = 0.61) showed unstable baselines with positive trends, while P9 (*Tau* = -0.43) showed an unstable baseline with a negative trend. Of the nine participants, both P1 (p = .013) and P3 (p = .004) demonstrated significant improvement during the intervention phase. Phase contrasts yielded a non-significant (p = .33) aggregated effect size of 0.10. P5 (*Tau*= 0.40) continued to show positive trend in the post-intervention maintenance phase, while P7 (*Tau*= -0.50), P8 (*Tau*= -0.40) and P9 (*Tau*= -0.40) showed negative trend.

Single subject treatment effects (grammaticality judgement). Pre-intervention baselines for past tense grammaticality judgement probes were stable for all participants. Only one participant (P5) improved significantly in correctly judging grammaticality on trained verbs tested within sessions (p = 0.02). P1 (p = 0.04) and P4 (p = 0.04) improved significantly on trained verbs tested between sessions, and a small (0.26) yet significant (p=.009) effect size across participants was calculated. Only one (P2) participant demonstrated significant trend in correct grammaticality judgement of untrained *-ed* verbs during the intervention phase (p = .02).

For grammaticality judgement 3S extension probes, P8 showed an unstable baseline with negative trend, *Tau*= -0.40. P4 demonstrated significant improvement during intervention (p = .02) and P8 demonstrated significant negative trend (p = .02). P2 (*Tau* = -0.80). Phase contrasts yielded a small, yet significant (p = .03) aggregated effect size of 0.22. P8 (*Tau* = -0.40) demonstrated negative trend in the maintenance phase, while P3 (*Tau* = 0.53) demonstrated positive trend.

For grammaticality judgement *'s* control probes, P4 demonstrated negative trend, while P7 (*Tau* =0.65) and P8 (*Tau* =0.90) demonstrated positive trend during baseline. P2 demonstrated significant positive trend during intervention (p = 0.02). Phase contrasts yielded a non-significant (p = .76) aggregated effect size of 0.03. P4 demonstrated negative trend in the maintenance phase, *Tau* = -0.40.

Within-group concurrent approach. Mean scores and standard deviations for *-ed* production and grammaticality judgement at four timepoints are presented in Table 3. A Friedman two-way ANOVA demonstrated that production of untrained *-ed* verbs differed significantly between timepoints, χ^2_F = 22.47, *df* =3, p <.001. Post-hoc Wilcoxon Signed Rank tests and a Bonferroni adjusted α of 0.0167 (0.05/3 comparisons: Timepoint 1 vs Timepoint 2; Timepoint 2 vs Timepoint 3, and; Timepoint 3 vs Timepoint 4) showed *-ed* production was significantly higher at Timepoint 3 (Mean Rank= 3.78) than at Timepoint 2 (Mean Rank= 1.56), z = -2.67, *N*-Ties = 9, p = .008. Differences between other Timepoints were non-significant, suggesting a stable pre-intervention baseline, an observable treatment effect between pre- and post-intervention testing points, and maintenance of gains at a group level. Tests for grammaticality judgement were non-significant.

Table 3

Mean scores on complete sets of untrained past tense verbs across four time points

	Pre-intervention		Post-intervention	
Measure	Timepoint 1	Timepoint 2	Timepoint 3	Timepoint 4
Expressive (/30)	7.44 (SD= 4)	7.44 (SD= 5.47) †	22.89 (SD= 5.97)*	21.89 (SD= 7.23) ††
Grammaticality judgement (/30)	15.22 (SD= 1.87)	16.22(SD= 1.03) †	19.25 (SD= 4.97)	18.78 (SD= 6.25)

Notes. SD= standard deviation.
†non-sig. difference between pre-intervention baseline timepoints= stable baseline.
*sig. difference between pre- and post-intervention timepoints= observed treatment effect.
††non-sig. difference between post-intervention timepoints= maintained treatment effect.

Table 4

Pre- and post-intervention standard scores

	SPELT-3		TROG-2	
Participant ID	Pre-intervention	Post-intervention	Pre-intervention	Post-intervention
P1	69	76 (2.78)*	74	76 (0.24)
P2	90	111 (9.33)*	97	95 (0.24)
P3	79	102 (6.83)*	86	93 (0.83)
P4	71	105 (13.54)*	81	83 (0.24)
P5	57	90 (13.14)*	81	86 (0.35)
P6	72	78 (0.64)	65	58 (-0.83)
P7	84	100 (6.37)*	62	74 (1.42)
P8	69	88 (7.54)*	79	97 (2.12)*
P9	57	78 (8.33)*	65	67 (0.24)

Notes. Scores are standard scores with a mean of 100 and SD of 15. RCI= reliable change index; SPELT-3= Structured Photographic Expressive Language Test 3rd Edition; TROG-2= Test of Reception of Grammar 2nd Edition.
*statistically significant, i.e. above 1.96.

Analysis of pre-post results. Pre- and post-intervention standard scores on the SPELT-3 and TROG-2 are reported in Table 4. Exceeding the RCI of 1.96 indicates statistically significant improvement. All but one participant (P6) exceeded the RCI for the SPELT-3. Further, for the majority of participants, post-intervention standard scores exceeded the manual-reported confidence intervals (90% and 95%) around their pre-intervention standard scores. Note, however, that even though P1's RCI was significant, his post-SPELT-3 standard score of 76 does not exceed the 90% and 95% confidence interval around his pre-SPELT-3 standard score of 69. One participant (P8) exceeded the RCI for the TROG-2 (2.12).

Adverse events

In the case of absence during the intervention phase, participants (P5, P6, P7, P8 and P9) attended a make-up session in the final week of intervention in which within session and between session teaching probes were collected. Due to issues with attention and engagement, procedural changes occurred for P6, who received 30 trials per session, and the systematic cueing hierarchy was limited to elicited imitation.

Discussion

This study evaluated the efficacy of an explicit grammar intervention combining metalinguistic training and grammar facilitation aimed to improve regular past tense (-*ed*) marking for nine children aged 5;10-6;8 years with DLD. Intervention taught -*ed* marking through explicit rule instruction and visual supports using the SHAPE CODING^{TM} system. A systematic cuing hierarchy (Smith-Lock et al., 2015) was used to support participants. This study contributes to the design, development and evaluation of intervention efficacy by moving through levels of evidence and analogous research designs (Robey, 2004).

Treatment effects

Single subject analyses. We hypothesised participants would improve significantly on *-ed* verbs trained and probed within sessions and between sessions. Most participants improved on expressive repeated measures of trained verbs with large effects, indicating this intervention is efficacious for improving production of *-ed* verbs taught in sessions. Further, most participants improved on untrained verbs with medium effects, suggesting generalisation. Within-group Friedman non-parametric two-way ANOVA also demonstrated a generalised treatment effect, which was maintained for five weeks. For grammaticality judgement, only three participants improved on trained verbs, one improved significantly on untrained verbs, and another continued to improve five weeks post-intervention. Few gains were observed across participants on an extension measure (3S) and on control measures of *'s* both production and grammaticality judgement. Limited progress on control probes strengthens our ability to attribute improvement on *-ed* production to intervention. Results support the efficacy of intervention to improve *-ed* production on trained and untrained verbs; however, we observed limited gains on grammaticality judgement measures.

Visual inspection of expressive repeated measures reflects results from statistical analysis regarding the immediacy of the functional relation between *-ed* production and intervention. That is, positive trend is observable upon the staggered introduction of intervention across participants. Specifically, trained expressive probes appeared to improve more rapidly, as early as week one of intervention, whereas for untrained verbs gains are observable around the five-week mark across participants. Finally, visual inspection revealed production of *-ed* on untrained verbs remained relatively stable for all children during the post-intervention phase, supporting findings from within-group statistical analysis.

Pre-post comparisons. Pre-post comparisons of standard measures of expressive and receptive grammar across participants mirrored single-subject analyses. Of the nine participants, eight exceeded the RCI for expressive grammar and one child exceeded the RCI for receptive grammar. Overall, pre-post analyses suggest the intervention had a broad effect on expressive grammar captured through standardised grammar measures. However, effects on measures of grammar comprehension were modest compared to expressive grammar.

General discussion

Results from the current study support and build upon findings in the literature. Finestack (2018) demonstrated efficacy of explicit-implicit instruction using novel morphemes, suggesting that the experimental approach may yield quicker gains, and improvement closer to mastery compared to existing implicit-only intervention procedures. Further, Finestack called for an evaluation of treatment effectiveness using true English morphemes across measures of maintenance and generalisation to progress the clinical applicability of research findings. Calder et al. (2018) piloted intervention with a small group of early school-age children diagnosed with DLD. Findings suggested intervention implemented over five weeks, twice per week without predefined dosage improved *-ed* production of untrained verbs and standard measures of expressive and receptive grammar. The authors concluded maintaining consistent dosage (i.e. 50 trials) and extending duration (i.e. 10 weeks) may improve production on untrained verbs and discern optimal dose to allow replication for clinical practice.

The current study applied recommended changes to intervention dose and intensity, and predictions were supported. Further, using measures of verbs trained in session and those from previous sessions allowed analysis of within- and between-session gains (e.g. Finestack, 2018). We saw that children showed greater and more rapid improvement on trained verbs probed within and between sessions compared to untrained verbs. However, gains in standard measures of receptive grammar were not observed to the extent reported in Calder et al. (2018). It is likely that reduced improvement on the measure is attributable to the baseline performance of the participants from the current study. That is, the baseline scores of the current group of participants were higher than those reported in Calder et al., which may suggest fewer gains were to be made on such a measure. This finding is consistent with literature suggesting that receptive grammar is less amenable to improvement when compared to expressive grammar (Ebbels, 2014).

From a theoretical perspective, limited improvement on receptive measures may be due to the status of internal representations of language remaining relatively fixed. However, increased production practice may establish new representations, such as those practised within sessions, which are generalizable to similar targets, such as other verbs marked for *-ed* or 3S. This pattern was observed with two participants (P2 and P4, respectively), so future research is needed explore this claim further. Alternatively, the current standard measures of receptive grammar may fall short of their aim. Recently, Frizelle et al. (2019) found multiple-choice grammar tasks may underestimate children's abilities compared to truth-value tasks. In the current study, probing grammaticality judgment of trained and untrained verbs allowed investigation of improvement of obligatory tense marking as a specific behavior, although improvement was limited across participants. This may provide evidence of the persistent nature of language disorder (e.g. Dale et al., 2018). Alternatively, the task may be implicated by other cognitive factors, such as phonological short-term memory. Regardless, further research is needed to unpack effective methods to treat receptive language difficulties.

Current findings are comparable to recent studies targeting *-ed* marking in children with DLD. For example, in a study using similar procedures to the current study, Smith-Lock et al. (2015) demonstrated explicit rule instruction coupled with a systematic cueing hierarchy was more effective in improving morphosyntax in preschool children with DLD when compared to recasting alone. A key difference to intervention procedures reported in this study is the inclusion of visual metalinguistic training and the *explicit* use of the cueing hierarchy. That is, cues in this study were presented to highlight the targeted behaviour was not observed, and so the children were encouraged to reflect on the rule they had just been taught with the support of visuals and to self-correct. Further, the current study implemented over double the cumulative intensity than Smith-Lock et al. (2015), although trials were not specified in that study, so it is challenging to make direct comparisons. Finally, Van Horne et al. (2017) reported positive treatment outcomes following intervention targeting *-ed* production. Importantly, the primarily implicit intervention procedures outlined in Van Horne et al. were effective in improving *-ed* for both studied groups following 36 sessions, which is markedly longer than dose duration reported here and by Smith-Lock et al. (2015), suggesting that explicit interventions may be more time efficient in improving expressive grammar outcomes. Future research is needed to compare the superiority of the two approaches to intervention.

This study further extends on a body of research evaluating the efficacy and effectiveness of explicit interventions using visual support strategies to improve grammatical knowledge for children with language difficulties, specifically, the SHAPE CODING™ system (Ebbels, 2007). Positive results of use of the system have been reported with older children with DLD (Ebbels et al., 2007, 2014; Kulkarni et al., 2013), younger children with DLD (Calder et al., 2018), and children with complex learning needs (Tobin & Ebbels, 2019). It should be noted that positive results were reported by Finestack (2018) where metalinguistic training without visual support was efficacious in improving grammar in young children with DLD. Continued research in this area will discern the extent to which the visual aspect of the SHAPE CODING™ system is responsible for positive treatment effects.

We saw that children showed greater and more rapid improvement on verbs trained in session when compared to untrained verbs, suggesting children with DLD may have difficulty generalizing grammar skills, particularly those relying upon sequence learning, such as finiteness marking. Therefore, we are more likely to see immediate improvement in verbs trained via intervention compared to untrained verbs. We also expected there might have been improvement on verbs marked for 3S, however this was not widespread across participants, with P6 improving during intervention, and three (P1, P2, P4) improving post-intervention. This finding suggests that, generally, grammar targets should be taught directly, even if they are linguistically related to existing intervention targets for children with DLD. Further, production practice did not seem to affect grammaticality judgment, however, metalinguistic training may have. That is, regardless of practice trials being held consistent, children for whom 'Silly Sentences' were introduced (P1, P2, P3, P4, P5, P7) appeared to perform better on repeated measures of grammaticality judgment (see S10, S11, S12). Therefore, introducing the sub-activity at the onset of treatment, rather than awaiting the 80% accuracy criterion, may result in improvement of grammaticality judgment.

Other factors to consider when evaluating treatment effectiveness are environmental. For example, the participant with the lowest performance in general (P6) had attended the specialist school for the least amount of time, compared to P2 and P4, the strongest performers who were in their third year at the specialist school. It could be that these children were primed to learn during language-based tasks more so than P6. However, P6 also had the lowest pre-intervention language scores and received fewer trials throughout the intervention phase. Nonetheless, P6 still improved significantly despite these potential barriers. Through SCEDs, evaluating individual treatment responses allows researchers and clinicians to extricate factors related to responsiveness to intervention that may otherwise be lost in group treatment studies (Plante, Tucci, Nicholas, Arizmendi, & Vance, 2018).

Limitations

There are limitations to this study. Firstly, generalizability of results using SCED must be applied with caution. Although the methodology allows for analysis of treatment effects for individuals, the lack of a control group and relatively small sample size inhibits the ability to make causal inferences regarding treatment effectiveness in relation to the general population. Further, within-participant analysis does not control for the influence of external factors, such as classroom instruction, when compared to robust randomized group comparison studies. Nonetheless, SCEDs provide a useful methodology for establishing an early evidence-base for newly developed interventions (Fey & Finestack, 2008). In fact, Horner et al. (2005) suggests results from a minimum of five studies totaling at least 20 participants across three different research teams are necessary to determine intervention efficacy using high quality SCEDs prior to effectiveness being tested using clinical trials. The current study was designed using guidelines developed by Kratochwill et al. (2012) and Tate et al. (2016) to meet minimum standards for SCED to interpret treatment efficacy. Note that an independent rater did not collect repeated measures within the baseline and intervention phases as per Kratochwill et al.'s (2012) recommendation. However, strong inter-rater reliability values addressed potential observer bias. Secondly, the current study used convenience sampling to recruit participants from a specialized school designed to provide intensive language and literacy support to young children with DLD. While non-verbal IQ was not directly measured as part of this study, all participants were enrolled into an educational program for children with DLD in the presence of average non-verbal IQ. Further, socio-economic status of participants was unknown and the majority (8/9) of participants were male. Therefore, the current sample may not be representative of the population of children with DLD at large. Lastly, the current efficacy study was limited to the analysis of *-ed* production and grammaticality judgment, and standard expressive and receptive grammar scores. More naturalistic measures, such as narrative or conversation sampling, may better serve as true measures of generalization in future studies.

Clinical implications

A recent survey of US speech pathologists investigating current clinical practices for grammar intervention found that although a regular component of practice, specific aspects of grammatical interventions are not well understood (Finestack & Satterlund, 2018). Further, -*ed* marking is often targeted as a treatment goal, and explicit presentation is often used in intervention procedures. However relatively little research has been reported using explicit intervention for teaching -*ed* to early school-aged children. Fey and Finestack (2008) proposed a framework for conceptualizing intervention components. The current intervention is summarized in Table 5. This framework may serve as a point of reference for clinicians planning to implement intervention to improve production of -*ed* for early school-aged children with DLD. Clear intervention procedures and maintaining consistent dose throughout the intervention phase also allows clinicians to replicate findings. It appears generally that this intervention is less efficacious for improving grammaticality judgment of -*ed*, with only a small intervention effect (0.26) observed. However, a similar effect (0.22) was observed for grammaticality judgement of 3S, but not for the production or grammaticality judgement of *'s*. Since 3S was not targeted directly but is linguistically related, perhaps improvement for some children was due to the phonological saliency of /z, s/ compared to -*ed* /d, t/ providing a learning advantage to the morpheme when combined with metalinguistic training.

Table 5

Framework for conceptualising intervention components proposed by Fey and Finestack (2008)

Intervention component	Experimental intervention
Children	5;10-6;8 year old children with DLD
Goals	Regular past tense (*-ed*) production and grammaticality judgment
Service delivery	1:1 with a speech-language pathologist in clinical contexts (within a specialized school)
Dosage	50 trials, 2x sessions per week for 10 weeks: 1000 trials over 20 sessions and ~7-10 hours of intervention
Procedures	Explicit intervention using metalinguistic training with visual support combined with an systematic cueing hierarchy
Activities	Naturalistic games with opportunities to produce *-ed* verbs (e.g. playdough, puppets, board games)
Measurement of outcomes	Standard grammar measures and criterion-referenced measures of *-ed* production and grammaticality judgment

Conclusions

Results continue to support the efficacy of explicit grammar interventions to improve -*ed* marking in early school-aged children. Future research should continue to evaluate the efficacy of similar interventions, for example, using more clinically relevant dosage (e.g. 1x session per week). It is also important to determine whether explicit grammar interventions can improve other aspects of grammatical difficulty for younger children with DLD, such as copula/auxiliary use, or *wh*- questions. Overall, findings contribute to the understanding of efficacious intervention procedures for early school-age children with DLD suggesting children are able to apply knowledge acquired through explicit instruction.

Appendix A

Checklist for scoring intervention procedure fidelity.

STEP	EXPLANATION	1/0
1. Explicit teaching	Remind child of the goal of the session	
1a. Activate prior knowledge		❒
1b. Explain Goals		❒
	ACTIVITY 1	
2. Check vocabulary	Child asked to label materials from session linked to subject/object nouns	❒
3. Goal	Demonstrate 3x SV/O sentences using one exemplar from each of the allomorphic categories. Introduce 'left down arrow cues' each alongside its corresponding shape	❒
4. Practice	25 trials to produce past tense *-ed* with systematic cueing	
4a. Coding	Lay large shapes on the floor and student to use as cues to produce SV/O sentences	❒
4b. Trials	22-28 trials achieved	❒
4c. Cueing	Errors cued appropriately?	❒
5. Consolidation	At the end of the session, review the 3x SV/O sentences using one exemplar from each of the allomorphic category.	
5a. Comprehension task	Student to produce SUBJECTs, VERBs, and OBJECTs following comprehension questions	❒
5b. Production	Student says phrase	❒
5c. Repeat without shapes	Student says phrase (cue as necessary)	❒
	ACTIVITY 2	
6. Check vocabulary	Child asked to label materials from session linked to subject/object nouns	❒
7. Goal	Demonstrate 3x SV/O sentences using one exemplar from each of the allomorphic categories. Introduce 'left down arrow cues' each alongside its corresponding shape	❒
8. Practice	25 trials to produce past tense *-ed* with systematic cueing	
8a. Coding	Lay large shapes on the floor and student to use as cues to produce SV/O sentences	❒
8b. Trials	22-28 trials achieved	❒
8c. Cueing	Errors cued appropriately?	❒
9. Consolidation	At the end of the session, review the 3x SV/O sentences using one exemplar from each of the allomorphic category.	
9a. Comprehension task	Student to produce SUBJECTs, VERBs, and OBJECTs following comprehension questions	❒
9b. Production	Student says phrase	❒
9c. Repeat without shapes	Student says phrase (cue as necessary)	❒
10. Summarise	Remind child of the goal of the session	❒
	TOTAL:	/19
	PERCENTAGE ACCURACY:	%

Appendix B

List of Supplemental Materials

S1: Expressive raw scores of participants on trained past tense verbs within-session.

S2: Expressive raw scores of participants on trained past tense verbs between-session.

S3: Expressive raw scores of participants on untrained past tense verbs.

S4: Expressive scores of participants on third person singular (extension).

S5: Summary of Tau-U analyses for expressive repeated measures baseline versus treatment phase contrasts on untrained third person singular targets (extension).

S6: Graph of % correct on expressive third person singular repeated measures (extension).

S7: Expressive raw scores of participants on possessive ‘s (control).

S8: Summary of expressive repeated measures baseline versus treatment phase contrasts on untrained possessive ‘s targets (control).

S9: Graph of % correct on expressive possessive *‘s* repeated measures (control).

S10: Grammaticality judgment raw scores of participants on trained past tense verbs within-session.

S11: Grammaticality judgment raw scores of participants on trained past tense verbs between-session.

S12: Grammaticality judgment raw scores of participants on untrained past tense verbs.

S13: Summary of grammaticality judgment repeated measures baseline versus treatment phase contrasts on trained and untrained targets.

S14: Graph of % correct on grammaticality judgment within-session repeated measures.

S15: Graph of % correct on grammaticality judgment between-session repeated measures.

S16: Graph of % correct on expressive untrained repeated measures.

S17: Grammaticality judgment raw scores of participants on third person singular (extension).

S18: Summary grammaticality judgment repeated measures baseline versus treatment phase contrasts on untrained third person singular targets (extension).

S19: Graph of % correct on grammaticality judgment third person singular repeated measures (extension).

S20: Grammaticality judgment raw scores of participants on possessive ‘s (control).

S21: Summary of grammaticality judgment repeated measures baseline versus treatment phase contrasts on untrained possessive ‘s targets (control).

S22: Graph of % correct on grammaticality judgment possessive ‘s repeated measures (control).