2024-11-01

Efficacy of Complexity-Based Target Selection for Treating Morphosyntactic Deficits in Children With Developmental Language Disorder and Children With Down Syndrome: A Single-Case Experimental Design

Kirsten M. Hannig Russell
Brigham Young University - Provo, kirsten.russell@byu.edu

Julie L. Wambaugh
University of Utah

John L. Davis
University of Texas at San Antonio

Sean M. Redmond
University of Utah

Original Publication Citation

Hannig Russell, K. M., Wambaugh, J. L., Davis, J. L., & Redmond, S. M. (2024). Efficacy of complexity-based target selection for treating morphosyntactic deficits in children with DLD and children with Down syndrome. American Journal of Speech Language Pathology, 33(2), 2939-2971. https://doi.org/10.1044/2024_AJSLP-24-00171

Efficacy of complexity-based target selection for treating morphosyntactic deficits in children with DLD and children with Down syndrome

Kirsten M. Hannig Russell[1], Julie L. Wambaugh[2], John L. Davis[3], Sean M. Redmond[2]

Brigham Young University[1], University of Utah[2], The University of Texas at San Antonio[3]

Author Note

Kirsten Hannig Russell https://orcid.org/0000-0002-6188-9494

Julie Wambaugh https://orcid.org/0000-0002-9888-0274

John Davis https://orcid.org/0000-0002-4394-2503

Sean Redmond https://orcid.org/0000-0002-3321-6022

Conflict of Interest

The authors have no relevant financial or nonfinancial conflicts of interest to disclose. This article is based on the dissertation completed by K.M.H. Russell (2023).

Funding

The research reported here was unfunded.

Correspondence

Correspondence concerning this article should be addressed to Kirsten Hannig Russell, 161 Taylor Building, Brigham Young University, Provo Utah, 84602. Email: kirsten.russell@byu.edu

Abstract

Purpose: Selecting targets for morphosyntactic intervention is a critical component of treatment planning. The complexity approach suggests that by treating a complex morphosyntactic target improvements will occur for the treated structure and for related, yet simpler, structures. This study evaluated the efficacy of the complexity approach for treating morphosyntactic deficits by targeting a complex BE verb question structure for children with DLD and children with Down syndrome (DS) and observing its impact on treated and untreated BE verb structures. We also considered whether etiology influenced participants' treatment responses.

Method: Three participants with DLD and three with DS received treatment for the BE verb question structure in the context of a multiple baseline design across participants. Accuracy of production for the treated structure and untreated BE verb structures was measured across baseline, treatment, and post-treatment phases.

Results: Treatment of the complex BE verb question structure resulted in change on the treated structure for three participants (i.e., two with DLD and one with DS). Generalization of treatment to untreated, related BE verb structures occurred for all six participants. Etiology did not appear to influence participants' responses to treatment.

Conclusions: This study provides evidence supporting the use of a complexity-based approach for selecting morphosyntactic treatment targets for children with DLD and children with DS. Additional research is needed to identify specific characteristics that may influence individual treatment responses.

Keywords: *developmental language disorder, Down syndrome, morphosyntax, intervention*

Efficacy of complexity-based target selection for treating morphosyntactic deficits in children with DLD and children with Down syndrome

Children with language disorders require intervention for their linguistic difficulties, yet traditional approaches for selecting morphosyntactic treatment targets require a large investment of time and only modest improvements (Law et al., 2004). The Complexity Account of Treatment Efficacy (CATE; Thompson et al., 2003) is a promising alternative to traditional selection methods. The CATE suggests training more complex morphosyntactic targets may generalize to untreated, simpler, yet linguistically related, targets. However, the efficacy of this approach for pediatric morphosyntactic intervention has not been established. Moreover, it may not be appropriate for all children with morphosyntactic deficits. Etiological differences could warrant different treatment target approaches (Bishop et al., 2017; Catts et al., 2002; Cole & Fey, 1996) or alternatively, have no impact on treatment outcomes (Cole et al., 1990). Developmental language disorder (DLD) and Down syndrome (DS) represent two conditions associated with morphosyntactic deficits that provide opportunity to examine these questions.

Current Approaches for Selecting Treatment Targets

Selecting treatment targets is likely more important than the therapeutic strategies utilized in intervention (Eisenberg, 2013). Current approaches for selecting morphosyntactic treatment targets include developmental or remedial approaches (Paul et al., 2018). The developmental approach selects treatment targets based on sequences observed in neurotypical development. It prioritizes emerging skills that fall within the child's zone of proximal development (ZPD; Vygotsky, 1978), or the space between what they can do independently and what they can do with adult support. Following neurotypical sequences ensures no stage of linguistic development is skipped. However, this approach likely exacerbates language learning gaps found between children with language disorders and their peers with neurotypical development (TD). Children with language disorders are already behind in their

linguistic development and treating each linguistic element in strict sequence nearly ensures they will not close the gap (Rice, 2020).

In contrast, the remedial approach to goal selection focuses on the most functional communication needs helping children function more effectively regardless of their morphosyntactic competencies. For example, an adolescent with DS may not be able to independently generate grammatically correct sentences to request, yet they may need to utilize public transportation to get from one place to another (e.g., school). Based on this need, a remedial approach would focus on script training to ensure they can tell the bus driver that they need to go to school. This approach supports communication but does not necessarily help clients independently generate similar utterances.

Outcomes using traditional approaches have been generally modest (d = 0.70, n = 271, CI = 0.14-1.55; Law et al., 2004). Because treatment targets may be one of the most critical aspects of effective intervention (Eisenberg, 2013), adjusting this factor could yield better outcomes.

The Complexity Approach

Young, TD children demonstrate emerging comprehension and production of grammatical morphemes and complex syntax simultaneously (Barako Arndt & Schuele, 2013). Consequently, targeting simpler, emerging linguistic skills before more complex skills may rest on a faulty developmental assumption. Incorporating "desirable difficulties" (Bjork, 1994) into treatment may increase efficacy, retention, and generalization. Indeed, many intervention studies have targeted absent morphosyntactic forms yet shown significant improvements (e.g., Camarata et al., 1994; Fey et al., 1993).

To show how a complexity approach to target selection may be applied in morphosyntactic treatment, let's consider the BE verb "*is*". Linguistic theory suggests that a canonical syntactic structure (e.g., *He is eating*) is simpler than a noncanonical syntactic structure (e.g., *Is he eating?*). In Figure 1, the canonical structure (Figure 1, Element A) is an independent clause in the form of a declarative sentence (hereafter *declarative*). Note that the words of the declarative are organized in the subject-verb-

participle order and include a Complementizer Phrase (CP) and an Inflectional Phrase (IP). The IP contains the tense of the clause. The Noun Phrase (NP; subject) is the specifier of the IP (e.g., "*He*"). The complement of the IP is the Verb Phrase (VP; predicate). The auxiliary verb "*is*" serves as the head of the VP and is inflected for tense and Number agreement with the NP. The lexical verb "*eating*" is the present participle and completes the VP. Note that the head of the phrase (e.g., *is*) has been moved to the IP prime level of the clause to reflect its tense features. Despite this movement, the order of the words remains in the Subject, BE verb, Verb+Present Participle order.

The noncanonical version "*Is he eating?*" is based on the canonical structure, but the word order has changed (Figure 1, Element B). Syntactic movement is responsible for this change, resulting in an interrogative (hereafter *question*) structure. To create a question structure, the auxiliary verb "*is*" must be moved to the head of the CP yet retains its tense and Number features. To understand the grammatical relationship between the auxiliary verb and the subject with which it agrees, one must recognize the underlying canonical structure while simultaneously recognizing the different surface structure of the question "*Is he eating?*". This movement and the necessity to link canonical and noncanonical structures (subject + predicate relationship) is why noncanonical structures are considered more complex than canonical structures.

De Anda et al. (2020) tested this idea of complexity in treatment using an AB single-case design in which they compared the accuracy of auxiliary BE question production by three participants before and after treatment. Results indicated treatment and improved production of auxiliary BE questions were associated. Improvements on related BE verb structures (copula and auxiliary sentences and copula questions) varied across participants. De Anda et al. suggested that the complexity approach may be an effective method for selecting treatment targets. However, the study design did not provide enough control to determine whether the treatment caused the outcomes observed.

Etiological considerations

Individual etiologies may likely influence, and possibly predict, individual responses to treatment using a complexity-based approach. Some researchers have posited that differences in cognitive abilities, such as those that exist between children with developmental language disorder (DLD) and children with Down syndrome (DS) may interact with syntactic processing requirements and require different interventions (Bishop et al., 2017; Catts et al., 2002; Cole & Fey, 1996).

Children with Developmental Language Disorder

DLD is a neurodevelopmental condition characterized by enduring deficits in one or more components of language (i.e., phonology, morphology, syntax, semantics, pragmatics) for 7%–11% of children (Norbury et al., 2016; Tomblin et al., 1997) who have no other primary conditions (e.g., Down syndrome, hearing impairment, intellectual disability, autism; Bishop et al., 2017). Nonverbal intelligence quotients (NIQs) can range from below average (but above thresholds for intellectual disability) to gifted levels of performance (e.g., standard scores of 70–130+; Bishop et al., 2017). A common profile associated with DLD consists of the presence of moderate-severe morphosyntactic deficits in the context of relatively stronger semantic skills (Rice et al., 2005). Morphosyntactic difficulties for English-speaking children with DLD typically involve omission errors with finite forms requiring tense and/or Number agreement (Bedore & Leonard, 1998). Omissions increase in sentence contexts requiring syntactic movement (Grela & Leonard, 2000; Rowland et al., 2005; van Der Lely, 1998).

Children with Down Syndrome

Down syndrome (DS) is the most common cause of intellectual disability and has a prevalence of 1 in 707 births (Mai et al., 2019). Approximately 80% of school-aged children with DS also have intellectual disability (IQ below a standard score of 70; Abbeduto et al., 2007). Children with DS who have language disorders demonstrate similar linguistic patterns to children with DLD. For example, children with DS demonstrate relatively better semantic abilities compared to their morphosyntactic skills (Abbeduto et al., 2007; Grela, 2002). These morphosyntactic deficits typically present as omission

errors impacting grammatical function words that inflect tense (e.g., BE verbs, past tense marking; Chapman et al., 1998; Eadie et al., 2002). Like children with DLD, omission errors increase for children with DS in sentence contexts requiring syntactic movement (Abbeduto et al., 2007; Grela, 2002).

The Present Study

The overarching goal of this study was to evaluate the efficacy of a complexity-based approach to treat morphosyntactic deficits for children with DLD and children with DS. This represents an extension of the ideas presented in De Anda et al. (2020). Specifically, we investigated the effect of training a more complex BE verb question structure (e.g., *Is the dog jumping?*) while observing its impact on untreated, less-complex auxiliary BE (e.g., *He is jumping*) and copula BE (e.g., *He is sad*) declaratives, and copula BE questions (e.g., *Is he sad?*). In this study we addressed three aims: (1) whether training the complex auxiliary BE question structure was causal to acquisition of that structure, (2) the extent to which training the auxiliary BE question structure resulted in generalization to untreated simpler, yet related auxiliary BE declarative, copula BE declarative, and copula BE question structures, and (3) whether treatment responses by children with DLD differed from children with DS.

Method

Approval for all aspects of the study was provided by the institutional review board at the University of Utah. Three students with DLD (ages 5–8) and three students with DS (ages 5–15) were recruited from advertisements shared through local school districts and Down syndrome foundations. To be considered eligible for the study, participants with DLD were required to have no history of neurological damage or diagnosis of language disorder secondary to a biomedical condition (Bishop et al., 2017). Participants with DS were required to have no co-occurring autism spectrum disorder, to ensure the data collected reflected the influence of DS only. All participants were required to be monolingual speakers of General American English, use oral language as their primary means of communication, be receiving services from an SLP, pass a hearing and phonological screening, and

present with morphosyntactic deficits. Additionally, all participants were required to demonstrate a mean length of utterance in words (MLUw) between 3.0 and 5.0 in a 50-utterance narrative language sample, reflecting an ability to produce an utterance long enough to accommodate an auxiliary BE question structure. MLUw values of 5.0 are associated with mastery levels of performance for auxiliary BE verbs so children who demonstrated an MLUw above 5.0 were excluded from this study. Additional requirements for inclusion into the study were below-criterion scores on the Test of Early Grammatical Impairment (TEGI; Rice & Wexler, 2001), and 60% or lower correct use in obligatory BE verb use on the TEGI Be/Do (Be) probe. To ensure we could observe possible generalization effects on the production of untreated BE forms (i.e., auxiliary declaratives, copular declaratives, and questions), participants were also required to score below 60% accuracy on at least one of those untreated forms.

Eligibility Measures

Narrative language samples were collected to facilitate the calculation of MLUw values. Fifty complete and intelligible utterances were taken from participants' combined retelling of two narratives: "Mr. Wuffles!" (Wiesner, 2013), and "A Porcupine Named Fluffy" (Lester & Munsinger, 2013/1986). Transcripts were coded and analyzed using the Systematic Analysis of Language Transcripts (SALT; Miller & Iglesias, 2019) conventions.

The Test of Early Grammatical Impairment (TEGI; Rice & Wexler, 2001) is a psychometrically robust assessment of children's ability to mark tense on a variety of verbs (Nitido & Plante, 2020). All participants who were aged 8;11 (years;months) or younger were required to demonstrate below-criterion performance on the Elicited Grammar Composite (EGC), reflecting general difficulty with marking tense in obligatory contexts. Participants with DS who were 9 years and older must have achieved an EGC below 93%—the highest criterion score for age 8;11. Additionally, participants were required to demonstrate difficulty using BE verbs through a criterion score below 60% on the TEGI Be/Do (Be) probe. Performance on untreated BE verb structures was calculated using the number of correct

productions divided by the total obligatory contexts for each untreated BE verb form from the TEGI Be/Do (Be) probe.

This study incorporated the use of the specific BE verb forms "*is"* and "*are"*. To confirm participants could produce recognizable /z/ and /r/ phonemes in word-final positions, they were required to pass a phonological screening that used words from Goldman-Fristoe Test of Articulation–Third Edition Sounds in Words subtest (e.g., *chee<u>se</u>, guit<u>ar</u>;* Goldman & Fristoe, 2015).

Descriptive Measures

Descriptive information (Table 1) was collected to help identify possible influences on participants' responses to treatment. The Peabody Picture Vocabulary Test, Fourth Edition (PPVT–4; Dunn & Dunn, 2007) was administered to account for the possible influence of vocabulary knowledge on outcomes. The Wechsler Nonverbal Scale of Ability (WNV; Wechsler & Naglieri, 2006), designed for clinical populations (including those with language deficits), was used to obtain information about the cognitive abilities of our participants.

Participants With DLD

The first participant with DLD (DLD P1) was a male aged 5;1 at entrance into the study. According to parental report, DLD P1 was evaluated at age 4 for concerns with his communication and was currently receiving 20-minute sessions weekly through the school. Participant 2 with DLD (DLD P2) was a female aged 6;11 who had been evaluated for communication concerns at age 2 and received early intervention services. She was currently receiving 30-minute sessions at school each week. Participant 3 with DLD (DLD P3) was a male aged 8;2. DLD P3 began receiving services beginning at age 2 and was currently receiving a combination of two private and two school-based sessions per week.

Participants With DS

DS participant 1 (DS P1) was a male with Down syndrome aged 12;6 at entrance into the study. He had previously received private and school-based services. He was currently receiving 60-90 minutes

of therapy at the local middle school each month. Participant 2 with DS (DS P2) was a male aged 12;8. He had received services beginning in elementary school and was now receiving services twice weekly at his local middle school. Participant 3 with DS (DS P3) was a female aged 7;3. She began receiving services at age 2. At the time of the study, she was receiving SLP services at school 20 minutes weekly.

Experimental Design

A single-subject nonconcurrent multiple baseline design across participants was used (Slocum et al., 2022). In single-subject design, each participant serves as their own control through the measurement of treated behaviors before (i.e., baseline) and during treatment (McReynolds & Kearns, 1983). To establish a functional relationship between treatment and outcomes, robust improvements on the treated behavior must be observed during the treatment phase across at least three participants (Kratochwill et al., 2010). In this study, the number of minimum baselines was determined a priori and balanced across the DLD and DS etiologies (i.e., DLD P1 and DS P1 = 5 baselines; DLD P2 and DS P2 = 7; DLD P3 and DS P3 = 9). Treatment began once participants (1) met their minimum number of pre-assigned baseline probes, and (2) demonstrated stable or non-ascending trend performance on the dependent variable (production of auxiliary BE questions). Adaptations were required for DLD P1 and DS P3 due to spontaneous improvements during the baseline phase. Following baseline collection, treatment phases began. Experimental probes measured children's performances on treated and untreated BE structures and were administered before treatment activities. Two treatment sessions were administered weekly. Each session provided 30 teaching episodes in the context of a story vignette. When participants demonstrated performance at previously established criteria on the treated behavior they moved to the post-treatment phase, which measured skill retention.

Experimental Stimuli

The primary dependent variable was the production of auxiliary BE questions on experimental probes. The secondary dependent variables were the production of copula BE questions and auxiliary

and copula BE declaratives in experimental probes. All probes consisted of stimuli using BE present tense verbs (e.g., *is, are*). One hundred lexical verbs and sixty modifiers, expected to be in the lexicon of children aged five years and younger (Fenson et al., 2007; Hall et al., 1984), were used. The selection of common, highly familiar, frequently used and early acquired words was deliberate to control for lexical complexity. Verbs used in auxiliary probes were counterbalanced across number of arguments. Additionally, verbs and modifiers in the probes were balanced across singular (*is*) and plural (*are*) conditions. Lexical items that began with /s/, /z/, or /r/ phonemes were excluded from the stimuli to avoid masking the targeted BE structure (e.g., *The girl is sleeping; The girl's sleeping; The girl sleeping*).

Probes elicited treated (Supplemental Material S1) and untreated structures (Supplemental Material S2) using picture stimuli and verbal prompts. For example, a picture of a jumping dog was presented with a verbal prompt following De Anda et al (2020): "*I wonder if the dog is jumping. Ask the puppet.*" Correct responses to the auxiliary BE question prompt included an auxiliary BE verb at the head of the question that agreed with a third-person subject, followed by a lexical verb with progressive *-ing* marking (e.g., "*Is the dog jumping?*", not "*Are the dog jumping?*"). As the grammatical structure was the target, participants could have used a subject or lexical verb different from what was expected and still receive credit for a correct production (e.g., "*Are they laughing?*" substituted for "*Are they crying?*"). Likewise, generalization probes for copula BE questions required a copula BE verb at the head of the utterance and a descriptive word at the end of the utterance (e.g., *Is the dog happy?*). The two declarative conditions required a third-person subject followed by an auxiliary or copula BE verb and a lexical verb or modifier at the end (e.g., *The dog is jumping*; *The dog is happy*). Re-prompting during probes occurred when the participant did not provide a response that provided obligatory context for a BE verb (e.g., *He likes dogs*). If the re-prompt still did not result in obligatory context for a BE verb, the response was considered unscorable and not calculated for that probe. In cases in which the participants may have responded to probe items with obligatory context for a different BE verb structure than

intended (e.g., provided a copula BE declarative structure when a copula BE question was expected), these responses were moved to the appropriate category (e.g., copula BE question structure) and the denominator of the total was adjusted.

Baseline Phase

Five baseline probes were set as the minimum based upon an a priori decision to follow the What Works Clearinghouse *Standards 1.0* (Kratochwill et al., 2010) and to use the Conservative Dual Criterion method for visual analysis (CDC; Fisher et al., 2003). Probes for auxiliary BE questions consisted of 10 prompts randomly selected from an established set of 100. Participant performance on untreated BE structures was also assessed at this time.

Treatment Phase

Once stable performance on baseline probes was established for the primary dependent variable, the treatment phase began. Experimental probes were used to monitor treatment progress and administered prior to each scheduled session, beginning with treatment session 2. Five different story vignettes were the format for intervention. The order of presented story vignettes was randomized per participant and sequences were repeated as needed. Participants received the story vignettes up to four times each (i.e., 20 maximum treatment sessions).

The decision for continuing or concluding treatment required a unique approach to ensure that we could fully evaluate the effect of complexity. As such, the criteria were based on: (a) probe performance on the treated structure, (b) performance on the untreated structures, and (c) the total number of treatment sessions delivered (Supplemental Material S3). Perceived change and mastery were two measures of change used to make this determination (Kallhoff & Wambaugh, 2021). Perceived change was indicated when at least one data point exceeded the highest baseline data point and exceeded the baseline mean by more than one standard deviation. Mastery was indicated when data points in the treatment phase reached 80% or higher across three of four consecutive probes. If

perceived change, but not mastery, was observed on the treated BE structure, treatment continued for up to 20 sessions. If perceived change on the treated BE structure was not observed across nine consecutive probes of the treatment phase, participants moved to the post-treatment phase. If mastery was observed on the treated BE structure and there was an upward trend for at least one of the generalization probes across the three most recent data points, treatment continued until the upward change plateaued or a maximum of 20 sessions was reached. Following treatment, data were collected at 2 and 6 weeks to determine whether improvements for treated and untreated items were retained.

Treatment Procedure

Two 15- to 30-minute treatment sessions were provided weekly over Zoom by a certified speech-language pathologist (first author). Treatment procedures followed established protocols: a script for each story vignette and a response chart to track cueing levels and participant responses (Supplemental Material S4). Story vignettes were designed following the procedures reported by De Anda et al. (2020) which were originally based on the assessment procedures of the TEGI (Rice & Wexler, 2001). A teaching episode was defined as the prompting and cues necessary to elicit a grammatically correct production of an auxiliary BE question from participants. Fifteen different verbs were used twice in each story vignette—once in a singular context and once in a plural context—for a total of 30 teaching episodes. Total session length was variable based on the amount of cueing that was required.

Parents helped participants sign on to their private Zoom session from a quiet place in their homes. Story vignettes incorporated two mice and a dog, who "completed" tasks relatable to the participant (e.g., getting ready for school). The participant's role was to ask the puppet questions about the mice and dog. The interventionist prompted for each teaching episode according to a scripted protocol. For example, the interventionist would show the dog brushing his teeth and say *"I wonder if the dog is brushing. Ask if the dog is brushing."* A graduated cueing hierarchy was used to ensure participants' success with teaching episodes and included three levels: (1) a repetition of the prompt, (2)

a model of the targeted response (e.g., "*I'll ask the puppet: is the mouse eating? Now you do it*"), and (3) an elicited imitation (e.g., "*Say, is the mouse eating.*"). The following incorrect responses initiated the cueing hierarchy: (a) no response, (b) no attempt of an auxiliary BE question, (c) use of the wrong auxiliary verb form (e.g., *is* for *are*), or (d) omission of the auxiliary BE form (e.g., *The dog eating?*).

Treatment Fidelity & Reliability

Fidelity of treatment delivery was ensured through pre-intervention training for the interventionist (first author) and the use of a script for each treatment session (Supplemental Material S4). Reliability of treatment delivery was evaluated for 10% of the treatment sessions per participant by a second observer. The number of correctly provided treatment components divided by the total number expected resulted in a treatment delivery reliability estimate of 100%. Point-to-point scoring reliability was completed for 20% of the administered probes per phase and per participant. A research assistant, blinded to participant status, scored the probes using session video recordings. Differences across the first and second scorer's results were marked as disagreements. The total number of agreements, divided by the combined number of agreements and disagreements, resulted in an estimated reliability of 87%.

Data Analysis Plan

Visual and statistical analyses of graphed data were used to determine the effect of treatment across participants. The Conservative Dual Criterion method (CDC; Fisher et al., 2003) was used for visual analysis of the primary dependent variable. This method utilizes baseline data to create level and trend lines, adjusted upward by 0.25 standard deviations and extended through the treatment phase. These lines provided conservative estimates of participants' expected performance on treated items if they did not receive treatment. If a pre-specified number of data points in the treatment phase fell above both CDC lines (see Fisher et al., 2003), a functional relationship between treatment and outcomes could be reliably established. Tau-U was the statistical measure used to calculate effect sizes, or the magnitude of

change between baseline and treatment phases (Parker et al., 2011). Tau-U was selected due to its ability to account for all data points within and across the baseline and treatment phases and to accommodate problematic baseline trends (Fingerhut et al., 2021). We elected to use a relatively conservative distribution for interpretation of effect sizes (Parker et al., 2011) where: (a) .93–1.00 = large effect, (b) .63–.92 = medium effect, and (c) 0–.62. = small effect. Tau-U and associated *p* values were calculated using the open-source online Tau calculator for single-case research (Vannest et al., 2016). A baseline corrected Tau was necessary to account for changes during the baseline phases for DLD P1 and DS P3. We used the online Baseline Corrected Tau Calculator (Tarlow, 2016) for these calculations.

To evaluate the extent to which treatment of auxiliary BE questions resulted in reliable generalization to the three untreated forms, demonstration of mastery (≥ 80%) or demonstration of clinically significant change (i.e., one or more data points in the treatment phase exceeding the baseline mean by two standard deviations) was required (c.f. Kallhoff & Wambaugh, 2021).

Results

Participants With DLD

DLD Participant 1

Production of Treated Auxiliary BE Questions

DLD P1 demonstrated performance variability (range 0%–50%) during the baseline phase (Figure 2). Assigned a priori to five baseline probes, the variability of his performance required 10 baseline probes to establish a descending trend across three probes before beginning treatment. DLD P1 received the maximum of 20 treatment sessions, as he did not meet the performance criterion to discontinue treatment earlier. DLD P1 received a cumulative intervention intensity of 20 doses, or 600 teaching episodes (i.e., 30 teaching episodes per dose/session). Upon application of treatment, an upward trend was observed and perceived change (> 42% accuracy) occurred for 16 of 19 treatment probes. Probe performance in the treatment phase did not meet CDC requirements for a reliable treatment effect, as

only one of 19 data points in the treatment phase fell above both CDC lines (13 were required). A baseline corrected Tau was necessary (Table 2), based on significant trend during the baseline phase (τ_{trendA} = .489, p = .05). Controlling for baseline trend, the overall effect size of treatment was small ($\tau_{Corrected}$ = .331, p = .14). Post-treatment probe performance was relatively stable, with 50%–54% accuracy rates on the 2- and 6-week follow-up probes.

Generalization to Untreated BE Structures

Generalization probes indicated treatment resulted in changes to the untreated BE auxiliary declarative structure, but not to the copula BE declarative or question structures (Figure 2). Baseline data for auxiliary BE declaratives ranged from 0%–73%. Clinically significant improvement above baseline levels (> 87%) was observed for four data points. Post-treatment probes indicated partial maintenance, with scores at 67% and 75% accuracy at 2- and 6-weeks post-treatment. Baseline data for copula BE questions ranged from 0%–60%. A high rate of variability was observed during the treatment phase, but the highest data point (50%) did not meet criteria for clinically significant change (> 68%). Post-treatment scores revealed a return to baseline levels. Baseline data for copula BE declaratives ranged from 0%–100%. DLD P1 achieved multiple scores of 100% accuracy on probes during the treatment phase, indicating mastery of this skill. However, due to the high baseline scores, clinically significant change could not be calculated. Post-treatment probes indicated variable retention (67% and 100% accuracy).

In summary, DLD P1 demonstrated a small treatment effect for the treated auxiliary BE question structure. Generalization to the untreated auxiliary BE declarative structure was observed but only partially maintained. Generalization to copula BE question and declarative structures was not observed.

DLD Participant 2

Production of Treated Auxiliary BE Questions

As shown in Figure 3, DLD P2 demonstrated a stable baseline with 0% accuracy on all seven probes assigned a priori. DLD P2 received 20 treatment sessions. Upon application of treatment, an

immediate change in accuracy rate was observed (range 30%–73%) and all 19 data points fell above both CDC lines. Tau-U (Table 2) indicated a large effect (τ_{AvsB} = 1.0, p < .001). Post-treatment probes indicated the skill was retained, with 50% and 60% accuracy at 2 and 6 weeks.

Generalization to Untreated BE Structures

Generalization probes indicated change for the untreated auxiliary and copula BE declarative structures and for the copula BE question structure (Figure 3). Baseline data for DLD P2 on auxiliary declaratives ranged from 40%–60%. Clinically significant improvement (> 76%) was observed during the treatment phase, indicating reliable generalization to the untreated auxiliary BE declarative structure. Post-treatment probes indicated a return to pre-treatment levels (28% and 60%). Baseline data for copula BE questions was consistent at 0% accuracy. There was a high rate of variability during the treatment phase (range 0%–67%) and clinically significant change was observed. Post-treatment probes indicated retention of this skill, with 67% and 50% accuracy at 2- and 6-weeks. Baseline data for the copula BE declaratives ranged from 29%–75%. Variability of performance continued throughout the treatment phase (range 0%–100%) and clinically significant change (> 89%) was observed. Post-treatment probes indicated declining retention (75% and 17% accuracy).

In summary, DLD P2 demonstrated a clear treatment effect for treated auxiliary BE questions, with generalization observed for all three untreated BE structures during treatment.

DLD Participant 3

Production of Treated Auxiliary BE Questions

DLD P3 demonstrated stability of performance (Figure 4) with 0% accuracy for each of the nine baseline probes assigned a priori. DLD P3 received the maximum of 20 treatment sessions. Upon treatment application, a near immediate change in performance occurred, with 50% accuracy demonstrated on the third probe. Performance trended upward (range 0%–80%), with 17 of 19 data

points falling above both CDC lines (13 were required). Tau-U (Table 2) indicated a medium effect (τ_{AvsB} = .894, $p < .001$). Post-treatment probes indicated declining retention (46% and 20% accuracy).

Generalization to Untreated BE Structures

Generalization probes (Figure 4) indicated changes occurred for the untreated auxiliary and copula BE declarative structures but not retained. Generalization to the copula BE question structure could not be determined. Baseline data for auxiliary BE declaratives ranged from 0%–50%. Clinically significant change (> 53%) was observed for two data points. Post-treatment probes indicated low but stable retention for these structures (43% and 44%). Baseline data for copula BE questions was nearly consistent at 0%. However, no attempt at the question structure was made for seven of the nine probes. For the data point that indicated 100%, DLD P3 attempted only one question structure, and it was produced accurately. Therefore, comparison of accuracy rates may not be appropriate to evaluate generalization of this structure for this participant. However, during the treatment phase, a noticeable increase in attempts and accuracy was observed. Post-treatment probes indicated a return to lower accuracy levels (0% and 20%). Baseline data for the copula BE declarative structures was variable (range 20%–90%). This variability continued into the treatment phase (range 0%–100%), and two probes indicated clinically significant change (> 94%). Post-treatment probes indicated retention (75% and 71%).

In summary, DLD P3 demonstrated a treatment effect for the treated auxiliary BE question structure and generalization to the untreated auxiliary and copula BE declarative structures. Generalization to the untreated copula BE question structure could not be determined due to limited attempts at the structure during baseline probes.

Participants With Down Syndrome

DS Participant 1

Production of Treated Auxiliary BE Questions

DS P1 demonstrated stability of performance during the baseline phase (Figure 5), with 0% accuracy on all five baseline probes, assigned a priori. DS P1 received the maximum of 20 treatment sessions. Upon application of treatment, a near immediate change was observed, with 50% accuracy on the second probe of the treatment phase. Despite variability (range 0%–50%), a small upward trend was observed as treatment continued. Fifteen of 19 data points fell above both CDC lines (13 data points were required). Tau-U (Table 2) indicated a medium effect (τ_{AvsB} = .790, p < .001). Post-treatment performance indicated retention of this skill, with scores at 25% and 38%.

Generalization to Untreated BE Structures

Generalization of treatment for DS P1 to untreated structures occurred for the copula BE question and auxiliary BE declarative structures, but not for the copula BE declarative structure (Figure 5). Baseline data for auxiliary BE declaratives ranged from 0%–38%. Clinically significant change (> 49%) was observed for two data points. Post-treatment probes indicated a lack of retention (33% and 29%). Baseline data for the copula BE question structure ranged from 0%–30%. Clinically significant change (> 50%) was observed for four data points. Post-treatment probes indicated retention (50% and 57%). Baseline data for copula BE declaratives was variable (range 17%–60%), which continued into the treatment phase. All data points fell below perceived change levels, indicating a lack of generalization of treatment to this BE structure. Post-treatment probes confirmed this lack of change (0% and 17%).

In summary, DS P1 demonstrated a clear treatment effect for the treated auxiliary BE question structure, with reliable generalization observed for copula BE questions and auxiliary BE declaratives. Generalization do copula BE declaratives was not observed.

DS Participant 2

Production of Treated Auxiliary BE Questions

DS P2 demonstrated stability of performance during baseline with 0% accuracy across all seven baseline probes assigned a priori (Figure 6). Upon application of treatment, a near immediate change in

accuracy was observed on the second probe, with 50% accuracy (one of two attempts was correct). However, this improvement was not maintained. Probe performance returned to 0% consistently throughout the remainder of the treatment phase. Only one of the 11 data points fell above both CDC lines (9 points were required). Due to low probe scores during the treatment phase, he was moved to the post-treatment phase after receiving 11 treatment sessions. Accordingly, he received a cumulative intervention intensity of 11 doses, or 330 teaching episodes. Tau-U (Table 2) indicated a small effect of treatment (τ_{AvsB} = .090, p = .75). Post-treatment probe performance was consistent with the treatment phase, with 14% and 0% accuracy on probes at 2 and 6 weeks.

Generalization to Untreated BE Structures

DS P2 demonstrated generalization of the treatment to the untreated auxiliary BE declarative structure and possible improvement to the copula BE question structure, but no change to the copula BE declarative structure (Figure 6). Baseline data for DS P2 on auxiliary BE declarative structures ranged from 23%–78%. Clinically significant change (> 86%) was observed on one treatment probe, indicating generalization of treatment to this untreated structure. Post-treatment probes indicated a return to pre-treatment levels, with scores at 50% and 57%. Baseline data for the copula BE questions were consistent (0% accuracy), which continued into the treatment phase. Post-treatment probes indicated a temporary improvement on this skill (40% accuracy) at 2 weeks, but this returned to 0% accuracy by 6 weeks post-treatment. Baseline data for copula BE declaratives were variable (range 33%–100%), continuing into the treatment phase. The maximum score during the treatment phase (80%) failed to indicate clinically significant change. Post-treatment probes indicated variable performance (60% and 43%).

In summary, DS P2 demonstrated a small, temporary effect of treatment for the treated auxiliary BE question structure. Generalization was observed for auxiliary BE declarative structure, but gains were not retained post-treatment. Generalization was not observed for untreated copula BE structures.

DS Participant 3

Production of Treated Auxiliary BE Questions

DS P3 demonstrated consistent performance during the baseline phase for eight of nine probes assigned a priori (Figure 7). However, on probe number nine, she demonstrated a sudden increase in accuracy, resulting in an additional eight probes, until it was determined that the baseline accuracy rate was not continuing to trend upward. Upon treatment application, an upward trend was observed, but all probe data points fell between the two CDC lines. Treatment was discontinued after nine treatment probes, as DS P3 met criteria to move to post-treatment early. DS P3 received a cumulative intervention intensity of nine doses, or 270 teaching episodes. Due to DS P3's positive baseline trend (τ_A = .419, p = .02), baseline corrected Tau-U was calculated ($\tau_{Corrected}$ = .118, p = .63) and indicated a small treatment effect (Table 2). Post-treatment probe performance was variable (33% and 63%).

Generalization to Untreated BE Structures

Generalization probes (Figure 7) indicated treatment generalized to the untreated copula BE question structure and but not the auxiliary and copula BE declarative structures. Baseline data for auxiliary BE declaratives ranged from 0%–67% accuracy. Clinically significant improvement (> 61%) was not observed. Post-treatment probes indicated consistent performance (41% and 40% accuracy). Baseline data for the copula BE questions was consistent at 0% initially, with variability beginning with the 11th baseline probe (range 25%-67%). This increase coincided with increased attempts at the structure. During the treatment phase, clinically significant change (> 66%) was observed on one data point. Post-treatment probes indicated her accuracy reduced (50% and 40%). Baseline probes for copula BE declaratives were variable (range 0%–75%) and this continued into the treatment phase (range 29%–63%). Clinically significant change (> 78%) was not observed. Post-treatment probes indicated stability of performance, with 60% accuracy at both probe sessions.

In summary, DS P3 demonstrated a small, nonsignificant treatment effect for the treated auxiliary BE question structure, with generalization to untreated copula BE questions during treatment.

Upward trends were observed for the untreated auxiliary and copula BE declarative structures, but they were not substantial enough to indicate reliable generalization effects of treatment.

Discussion

The goal of this study was to evaluate the efficacy of a complexity-based approach to morphosyntactic treatment for school-aged children with DLD and school-aged children with Down syndrome. We utilized a multiple baseline design across participants to measure the effect of training a complex auxiliary BE question structure while observing the impact on untreated, less-complex BE structures (i.e., auxiliary BE declarative, copula BE question and declarative). We also sought to determine whether children with DLD and children with DS demonstrated differential treatment responses. Overall, the results support the efficacy of this approach for improvement of the trained structure and generalization to untreated structures. Etiology did not appear to influence treatment responses among participants.

Efficacy of Treating Auxiliary BE Question Structure

Three of the six participants—two with DLD and one with DS—demonstrated a reliable treatment effect for the treated auxiliary BE question structure, providing moderate evidence (Kratochwill et al., 2010) that the treatment was efficacious. However, none of the participants demonstrated acquisition (i.e., mastery). It is likely that there are conditions associated with how the complexity approach impacts the linguistic systems of those receiving treatment. For example, maintaining complexity might be needed to ensure ongoing progression. When progress on the original target occurs, adjustments to more challenging complex structures (e.g. adding WH-movement or more clauses into targeted question forms) may be required for individuals to progress toward mastery.

Identifying predictors of positive treatment effects represents an important aspect of establishing the validity of new treatment approaches. Examination of cognitive and linguistic characteristics of our participants did not reveal clear predictors differentiating treatment responders

from non-responders. Our data suggest stable 0% baselines may be associated with positive treatment effects (see data for DLD P2, DLD P3, and DS P1). However, stable 0% baselines cannot be the sole predictor (see data for DS P2). Etiology cannot explain this difference either; those who demonstrated treatment effects included both DLD and DS participants.

Another possible explanation for differences in treatment response may be related to dose. Four of the participants received 600 teaching episodes across 20 sessions whereas DS P2 received only 330 teaching episodes across 11 sessions. It is possible that DS P2 needed more therapeutic exposures to the treated structure before observable changes in his linguistic system could occur.

The two participants (DS P3 and DLD P1) who demonstrated spontaneous growth for auxiliary BE questions during baseline require additional consideration. The variability demonstrated during each of their baseline phases suggested the likelihood that the auxiliary BE question structure was in their zone of proximal development, and therefore, not truly a complex structure for them. Exposure to the probes likely assisted them in discerning key elements of that structure without treatment. Individual characteristics did not reveal similarities that could otherwise explain this spontaneous growth.

Generalization to Untreated Structures

The complexity approach is fully examined only when we consider the impact of treatment on related, simpler structures. To establish a functional relationship between treatment and generalization to untreated structures, three or more participants needed to demonstrate clinically significant change on untreated auxiliary and/or copula BE generalization structures. All three participants who sh reliable change on the treated BE structure also demonstrated clinically significant change in at least two of three untreated BE verb structures. It is important to highlight that all participants—even those who did not demonstrate treatment effects on the treated auxiliary BE question structure—demonstrated clinically significant change with at least one of the untreated structures during the treatment phase. Although the ideal of achieving mastery on treated and untreated structures was not realized in this

study, the efficacy of the complexity approach to treat not only one target but also effect change in related simpler morphosyntactic BE verb structures was still supported by our data.

Two of the three participants who demonstrated generalization to the copula BE question structure retained their progress. The principles of the complexity approach suggest that these participants acquired the underlying syntactic movement for BE question structures from treatment and successfully applied it to copula BE question structures. DS P3's lack of retention may have been related to the timing of the clinically significant change. DLD P2 and DS P1 demonstrated change for several consecutive probes during treatment, while DS P3 demonstrated this level of change only on her last treatment probe. An extension of treatment would likely have increased her retention.

Clinical Implications

Patterns of outcomes across the participants indicated that those who experienced the greatest improvement on the treated, complex structure were also those who showed the greatest generalization effects. This aligns with the complexity approach because, theoretically, key elements of a complex structure are extracted during treatment and available for simpler, related structures, resulting in broad changes to the linguistic system. However, generalization effects were observed across all participants, suggesting that using complexity to guide target selection supports progress on a range of morphosyntactic structures, regardless of whether treated structures improve.

Limitations

This study was designed to evaluate the efficacy of a complexity approach for treatment and generalization of BE verbs by capitalizing on syntactic movement. It was not designed to evaluate whether complexity was more efficient compared to other approaches. Although this approach is efficacious for treatment of BE forms, our results cannot be generalized to all morphosyntactic treatment targets. Subsequent investigations should evaluate the contributions of various elements of

morphosyntactic complexity (e.g., movement, number of grammatical features) to determine the role complexity plays in the development of grammatical systems.

Although the inclusion of participants from two different etiologies (i.e., DLD and DS) was a relative strength in our design, this represents only a small fraction of the variability present in the cognitive-linguistic profiles of children with language disorders. A full consideration of the value of the complexity approach requires examining treatment effects in other neurodevelopmental conditions.

Our treatment used implicit instruction; utilizing a cueing hierarchy to support participants in learning the auxiliary BE question structure. While implicit instruction is a valid method for treatment of morphosyntax (Eidsvåg et al., 2019), explicit instruction may have enhanced or improved acquisition effects for our participants (Finestack, 2018). Future studies should compare the differential effects of providing either explicit or implicit instruction during treatment of complex morphosyntactic targets.

Future Research

This study requires replication to confirm and expand our findings. Although our findings indicated that etiology did not impact participants' responses to treatment, the lack of observable patterns makes it difficult to determine who might benefit most from a complexity-based approach.

Conclusion

This study represents an important first step in establishing a causal relationship between the complexity approach for BE verb treatment and acquisition and generalization effects for school-aged children with DLD and school-aged children with DS. Differences in etiology did not appear to be a factor in predicting participants' responses to treatment. Additional research is needed to identify a range of efficacious complex morphosyntactic targets, and the individuals for whom this approach would be most appropriate.

Acknowledgments

We gratefully acknowledge the contribution of the participants and their families to this investigation.

Data Availability Statement

Data collected for the current study are available upon reasonable request.

References

Abbeduto, L., Warren, S. F., & Conners, F. A. (2007). Language development in Down syndrome: From the prelinguistic period to the acquisition of literacy. *Mental Retardation and Developmental Disabilities Research Reviews, 13*(3), 247–261. https://doi.org/10.1002/mrdd.20158

Barako Arndt, K. B., & Schuele, C. M. (2013). Multiclausal utterances aren't just for big kids: A framework for analysis of complex syntax production in spoken language of preschool-and early school-age children. *Topics in language disorders*, *33*(2), 125-139. https://doi.org/10.1097/TLD.0b013e31828f9ee8

Bedore, L. M., & Leonard, L. B. (1998). Specific language impairment and grammatical morphology: A discriminant function analysis. *Journal of Speech, Language, and Hearing Research, 41*(5), 1185–1192. https://doi.org/10.1044/jslhr.4105.1185

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., Adams, C., Archibald, L., Baird, G., Bauer, A., Bellair, J., Boyle, C., Brownlie, E., Carter, G., Clark, B., Clegg, J., Cohen, N., Conti-Ramsden, G., Dockrell, J., Dunn, J., Ebbels, S., … Grist, M. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry, 58*(10), 1068–1080. https://doi.org/10.1111/jcpp.12721

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe, & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). MIT Press.

Camarata, S. M., Nelson, K. E., & Camarata, M. N. (1994). Comparison of conversational recasting and imitative procedures for training grammatical structures in children with specific language

impairment. *Journal of Speech and Hearing Research, 37*(6), 1414–1423. https://doi.org/10.1044/jshr.3706.1414

Catts, H. W., Fey, M. E., Tomblin, J. B., & Zhang, X. (2002). A longitudinal investigation of reading outcomes in children with language impairments. *Journal of Speech, Language, and Hearing Research, 45*(6), 1142–1157. https://doi.org/10.1044/1092-4388(2002/093)

Chapman, R. S., Seung, H. K., Schwartz, S. E., & Kay-Raining Bird, E. (1998). Language skills of children and adolescents with Down syndrome: II. Production deficits. *Journal of Speech, Language, and Hearing Research, 41*(4), 861–873. https://doi.org/10.1044/jslhr.4104.861

Cole, K. N., & Fey, M. E. (1996). Cognitive referencing in language assessment. In K. N. Cole, P. S. Dale, & D. J. Thal (Eds.), *Assessment of communication and language* (Vol. 6, pp. 143–160). Paul H. Brookes Publishing Co.

Cole, K. N., Dale, P. S., & Mills, P. E. (1990). Defining language delay in young children by cognitive referencing: Are we saying more than we know? *Applied Psycholinguistics*, *11*(3), 291–302. https://doi.org/10.1017/S0142716400008900

De Anda, S., Blossom, M., & Abel, A. D. (2020). Cross-morpheme generalization using a complexity approach in school-age children. *Journal of Speech, Language, and Hearing Research, 63*(10), 3501–3524. https://doi.org/10.1044/2020_JSLHR-19-00173

Dunn, L., & Dunn, L. (2007). *Peabody picture vocabulary test* (4th ed.). Pearson Assessments.

Eadie, P. A., Fey, M. E., Douglas, J. M., & Parsons, C. L. (2002). Profiles of grammatical morphology and sentence imitation in children with specific language impairment and Down syndrome. *Journal of Speech, Language, and Hearing Research, 45*(4), 720–732. https://doi.org/10.1044/1092-4388(2002/058)

Eidsvåg, S. S., Plante, E., Oglivie, T., Privette, C., & Mailend, M. L. (2019). Individual versus small group treatment of morphological errors for children with developmental language disorder. *Language,*

Speech, and Hearing Services in Schools, *50*(2), 237–252. https://doi.org/10.1044/2018_LSHSS-18-0033

Eisenberg, S. L. (2013). Grammar intervention: Content and procedures for facilitating children's language development. *Topics in Language Disorders, 33*(2), 165-178. https://doi.org/10.1097/TLD.0b013e31828ef28e

Fenson, L., Marchman, V., Thal, D.J., Dale, P.S., Reznick, J.S., & Bates, E. (2007). *Communicative development inventories.* Brookes.

Fey, M. E., Cleave, P. L., 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*(1), 141–157. https://doi.org/10.1044/jshr.3601.141

Finestack, L. H. (2018). Evaluation of an explicit intervention to teach novel grammatical forms to children with developmental language disorder. *Journal of Speech, Language, and Hearing Research*, *61*(8), 2062–2075. https://doi.org/10.1044/2018_JSLHR-L-17-0339

Fingerhut, J., Xu, X., & Moeyaert, M. (2021). Impact of within-case variability on Tau- U indices and the hierarchical linear modeling approach for multiple-baseline design data: A Monte Carlo simulation study. *Evidence Based Communication Assessment and Intervention, 15*(3), 115–141. https://doi.org/10.1080/17489539.2021.1933727

Fisher, W. W., Kelley, M. E., & Lomas, J. E. (2003). Visual aids and structured criteria for improving visual inspection and interpretation of single-case designs. *Journal of Applied Behavior Analysis, 36*, 387–406. https://doi.org/10.1901/jaba.2003.36- 387

Goldman, R., & Fristoe, M. (2015). *Goldman-Fristoe test of articulation* (3rd ed.). Pearson.

Grela, B. G. (2002). Lexical verb diversity in children with Down syndrome. *Clinical Linguistics & Phonetics, 16*(4), 251–263. https://doi.org/10.1080/02699200210131987

Grela, B. G., & Leonard, L. B. (2000). The influence of argument-structure complexity on the use of auxiliary verbs by children with SLI. *Journal of Speech, Language, and Hearing Research, 43*(5), 1115–1125. https://doi.org/10.1044/jslhr.4305.1115

Hall, W. S., Nagy, W. E., Linn, R. L., & Bruce, B. (1984). *Spoken words, effects of situation and social group on oral word usage and frequency.* Erlbaum.

Kallhoff, L., & Wambaugh, J. (2021). Generalization outcomes of Sound Production Treatment (SPT) for acquired apraxia of speech: Relationship to treatment dosage. *Aphasiology, 37*(3), 382–407. https://doi.org/10.1080/02687038.2021.2020715

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M & Shadish, W. R. (2010). Single-case designs technical documentation. *What Works Clearinghouse*. https://files.eric.ed.gov/fulltext/ED510743.pdf

Law, J., Garrett, Z., & Nye, C. (2004). The efficacy of treatment for children with developmental speech and language delay/disorder: A meta-analysis. *Journal of Speech, Language, and Hearing Research, 47*(4), 924–943. https://doi.org/10.1044/1092-4388(2004/069)

Lester, H., & Munsinger, L. (2013/1986). *A porcupine named Fluffy.* Houghton Mifflin Harcourt.

Mai, C. T., Isenburg, J. L., Canfield, M. A., Meyer, R. E., Correa, A., Alverson, C. J., Lupo, P.J., Riehle-Colarusso, T., Cho, S. J., Aggarwal, D., Kirby, R. S., & National Birth Defects Prevention Network (2019). National population-based estimates for major birth defects, 2010-2014. *Birth Defects Research, 111*(18), 1420–1435. https://doi.org/10.1002/bdr2.1589

McReynolds, L. V., & Kearns, K. P. (1983). *Single-subject experimental designs in communicative disorders*. University Park Press.

Miller, J., & Iglesias, A. (2019). *Systematic Analysis of Language Transcripts* (SALT; Clinical Version 20) [Computer software]. SALT Software, LLC. https://www.saltsoftware.com/products/software

Nitido, H., & Plante, E. (2020). Diagnosis of developmental language disorder in research studies. *Journal of Speech, Language, and Hearing Research*, *63*(8), 2777-2788. https://doi.org/10.1044/2020_JSLHR-20-00091

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. https://doi.org/10.1111/jcpp.12573

Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining nonoverlap and trend for single-case research: Tau-U. *Behavior Therapy, 42*(2), 284–299. https://doi.org/10.1016/j.beth.2010.08.006

Paul, R., Norbury, C., & Gosse, C. (2018). *Language disorders from infancy through adolescence: Listening, speaking, reading, writing, and communicating* (5th ed.). Mosby.

Rice, M. L. (2020). Clinical lessons from studies of children with specific language impairment. *Perspectives of the ASHA Special Interest Groups, 5*(1), 12–29. https://doi.org/10.1044/2019_PERSP-19-00011

Rice, M. L., & Wexler, K. (2001). *Test of early grammatical impairment.* The Psychological Corporation

Rice, M. L., Warren, S. F., & Betz, S. K. (2005). Language symptoms of developmental language disorders: An overview of autism, Down syndrome, fragile X, specific language impairment, and Williams syndrome. *Applied Psycholinguistics, 26*(1), 7–27. https://doi.org/10.1017/S0142716405050034

Rowland, C. F., Pine, J. M., Lieven, E. V. M., & Theakston, A. L. (2005). The incidence of error in young children's wh-questions. *Journal of Speech, Language, and Hearing Research, 48*(2), 384–404. https://doi.org/10.1044/1092-4388(2005/027)

Russell, K. M. H. (2023). *Evaluation of a complexity-based morphosyntactic treatment for children with language disorders of mixed etiologies* (Publication No. 30811790) [Doctoral dissertation, University of Utah]. ProQuest Dissertations & Theses Global.

Slocum, T. A., Pinkelman, S. E., Joslyn, P. R., & Nichols, B. (2022). Threats to internal validity in multiple-baseline design variations. *Perspectives on Behavior Science*, *45*(3), 619-638. https://doi.org/10.1007/S30614-022-00326-1

Tarlow, K. R. (2016). Baseline corrected Tau calculator. http://www.ktarlow.com/stats/tau

Thompson, C. K., Shapiro, L. P., Kiran, S., & Sobecks, J. (2003). The role of syntactic complexity in treatment of sentence deficits in agrammatic aphasia: The complexity account of treatment efficacy (CATE). *Journal of Speech, Language, and Hearing Research, 46*(3), 591–607. https://doi.org/10.1044/1092- 4388(2003/047)

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research, 40*(6), 1245–1260. https://doi.org/10.1044/jslhr.4006.1245

van Der Lely, H. (1998). SLI children: Movement, economy, and deficits in the computational-syntactic system, *Language Acquisition, 7*, 161-192. https://doi.org/10.1207/s15327817la0702-4_4

Vannest, K. J., Parker, R. I., Gonen, O., & Adiguzel, T. (2016). *Single Case Research: web based calculators for SCR analysis. (Version 2.0)* [Web-based application]. Texas A&M University. singlecaseresearch.org

Vygotsky, L. S. (1978). *Mind in society*. Harvard University Press.

Wechsler, D., & Naglieri, J. A. (2006). *Wechsler nonverbal scale of ability*. Pearson.

Wiesner, D. (2013). *Mr. Wuffles*! Houghton Mifflin Harcourt.

Table 1.

Participant Characteristics

	Children with DLD			Children with Down syndrome		
	Participant 1	Participant 2	Participant 3	Participant 1	Participant 2	Participant 3
Age	5;1	6;11	8;2	12;6	12;8	7;3
MLUw	3.88	4.56	4.92	3.74	4.44	3.02
TEGI EGC[a]	44	51	29	7.5	34	15
Be/Do (BE) probe[a]	13	50	38	0	25	29
PPVT[b]	114	86	69	48	26	60
WNV[b]	87	120	91	41	39	71

Note. Age reported as years;months. MLUw = Mean Length of Utterance in words. TEGI EGC = Test of Early Grammatical Impairment Elicited Grammar Composite. PPVT = Peabody Picture Vocabulary Test, Fourth Edition. WNV = Wechsler Nonverbal Scales of Ability Full Scale Score.

[a]Criterion score.

[b]Standard score.

Table 2.

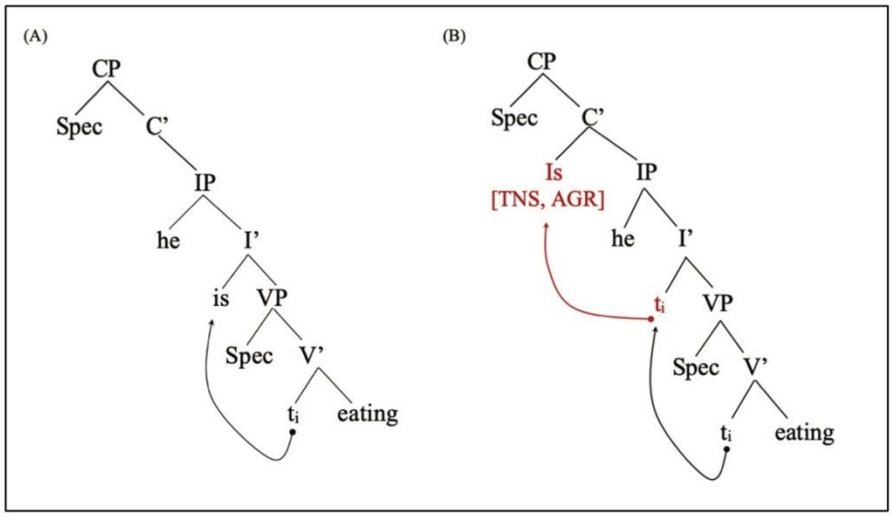
Tau-U Calculations Within and Across Phases per Participant

Participant	Variable	*S*	# pairs	Tau-U	*p*-value
DLD P1	Tau-UA vs B	85	190	0.447	0.05
	Tau-UCorrected	63	190	0.331	0.14
DLD P2	Tau-UA vs B	133	133	1	<0.0001
DLD P3	Tau-UA vs B	153	171	0.894	0.0002
DS P1	Tau-UA vs B	75	95	0.790	0.008
DS P2	Tau-UA vs B	7	77	0.090	0.75
DS P3	Tau-UA vs B	75	153	0.490	0.04
	Tau-UCorrected	18	153	0.118	0.63

Note. Tau-UA vs B indicates the degree of overlapping data points between the baseline and treatment phases. Tau-UCorrected incorporates a baseline corrected Tau-U estimate between the baseline and treatment phases where baseline data demonstrated a positive trend.

Figure 1.

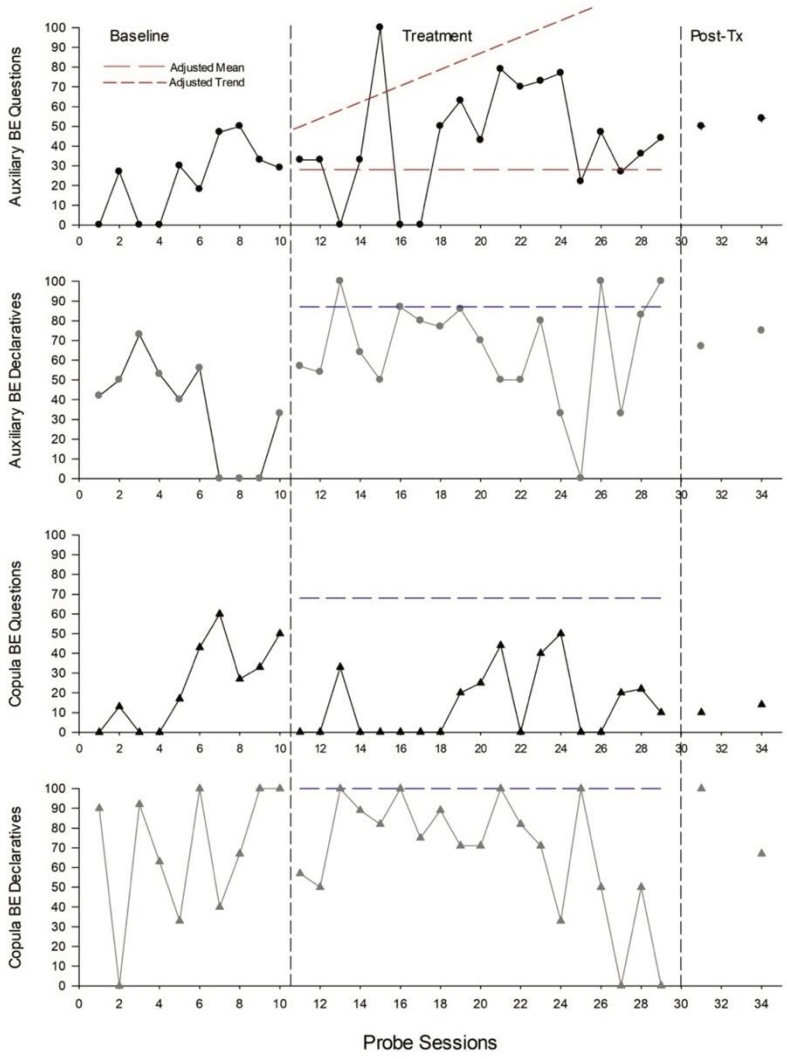
Syntax Trees Showing Required Movement for Auxiliary BE Question Structure.



Note. A = Auxiliary BE declarative structure. B = Auxiliary BE question structure showing an additional level of syntactic movement necessary to create the question form from the declarative (A) form.

Figure 2.

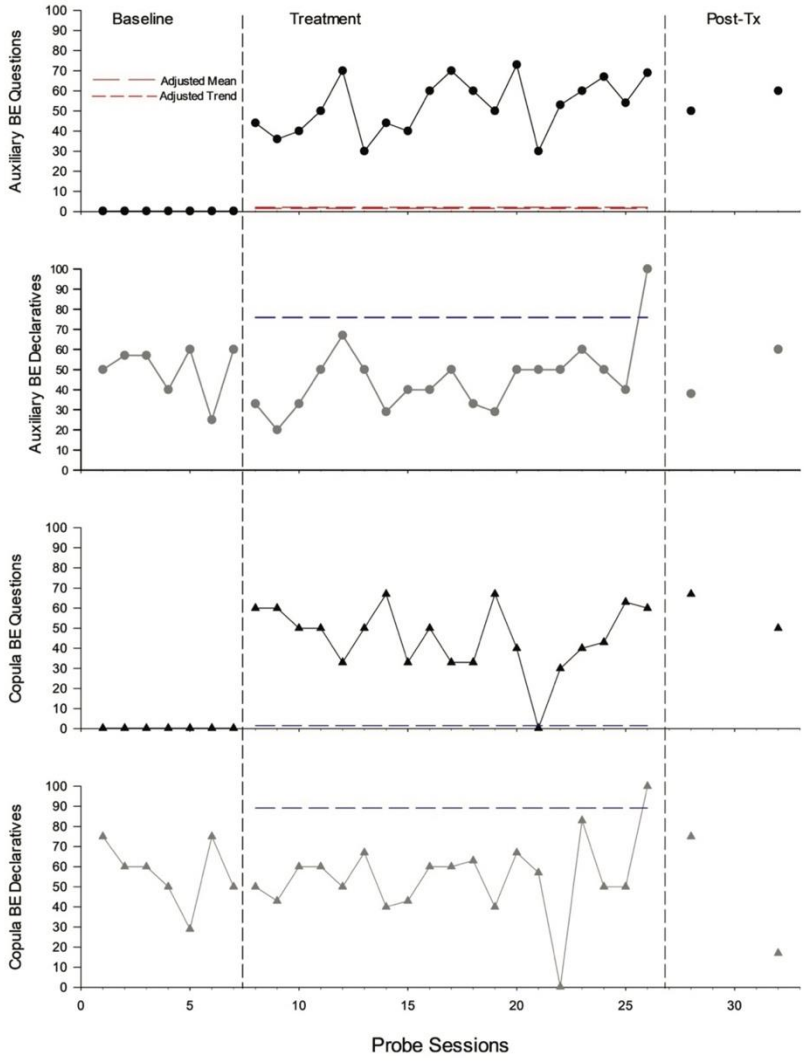
Percent Correct Auxiliary BE Questions and Untreated BE Structures on Probes for DLD P1.



Note. Dashed black lines indicate clinically significant change for the generalization item. Post-Tx = probes administered at 2- and 6-weeks post-treatment.

Figure 3.

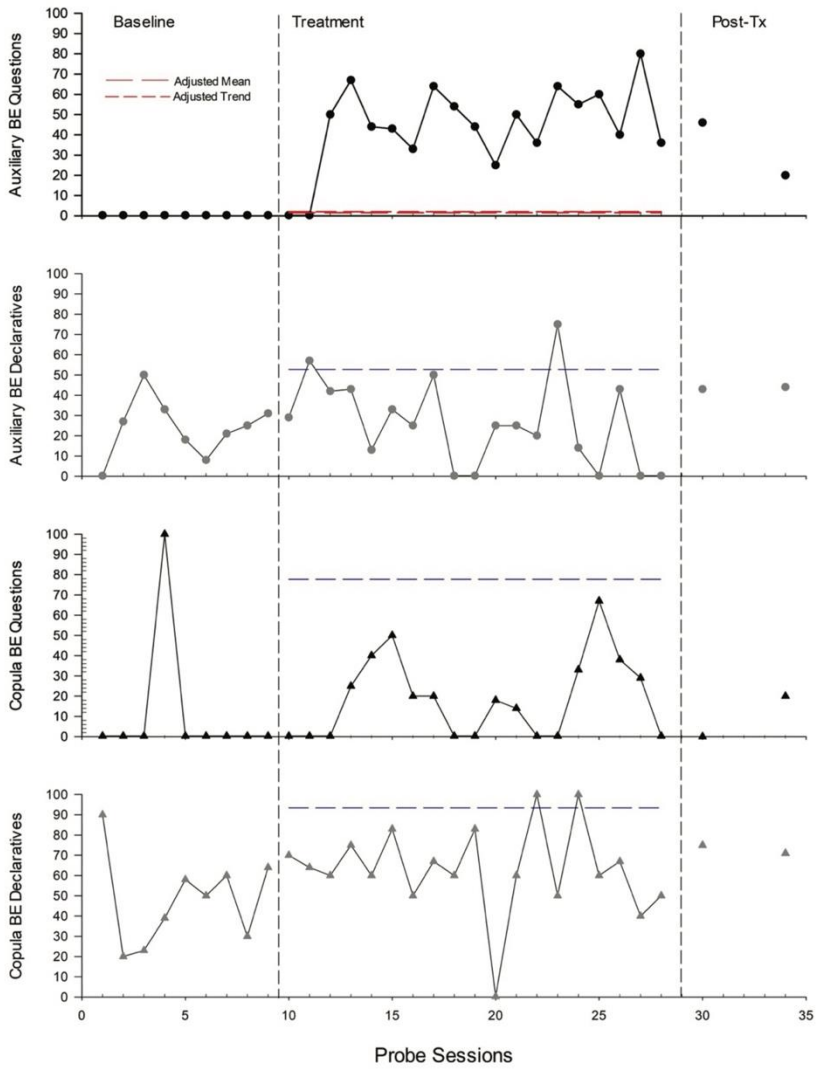
Percent Correct Auxiliary BE Questions and Untreated BE Structures on Probes for DLD P2



Note Dashed black lines indicate clinically significant change for the generalization item. Post-Tx = probes administered at 2- and 6-weeks post-treatment.

Figure 4.

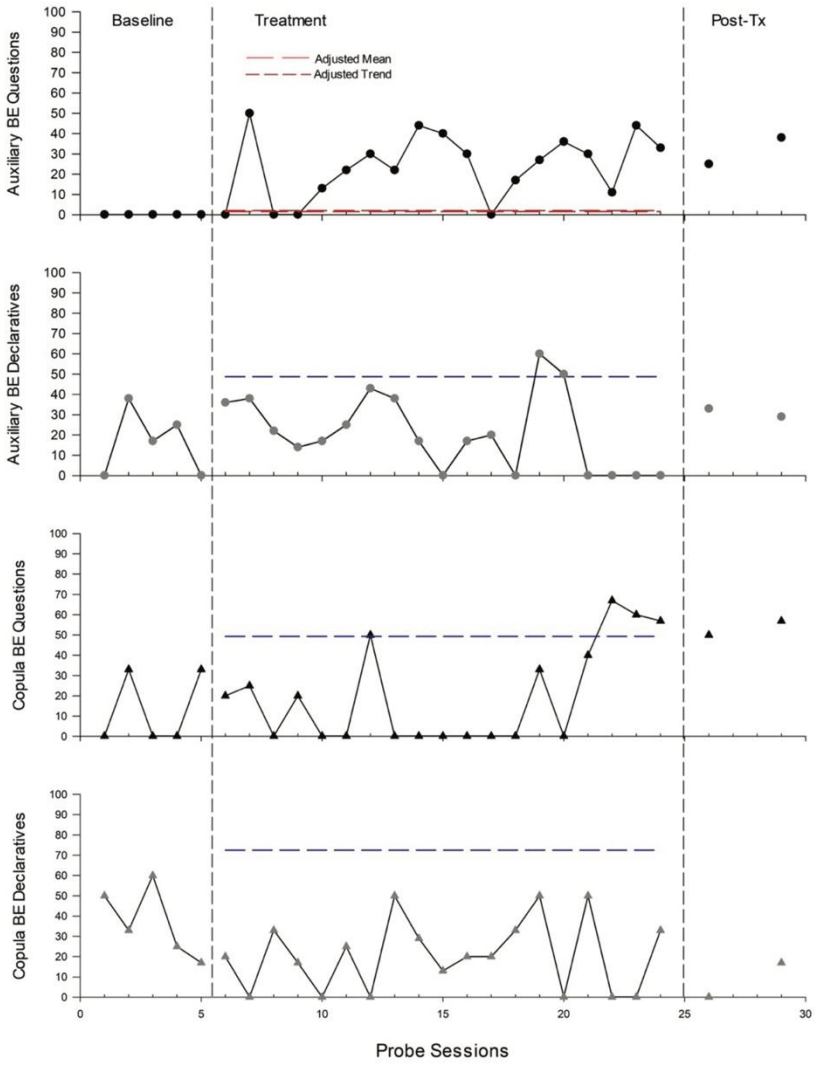
Percent Correct Auxiliary BE Questions and Untreated BE Structures on Probes for DLD P3



Note Dashed black lines indicate clinically significant change for the generalization item. Post-Tx = probes administered at 2- and 6-weeks post-treatment.

Figure 5.

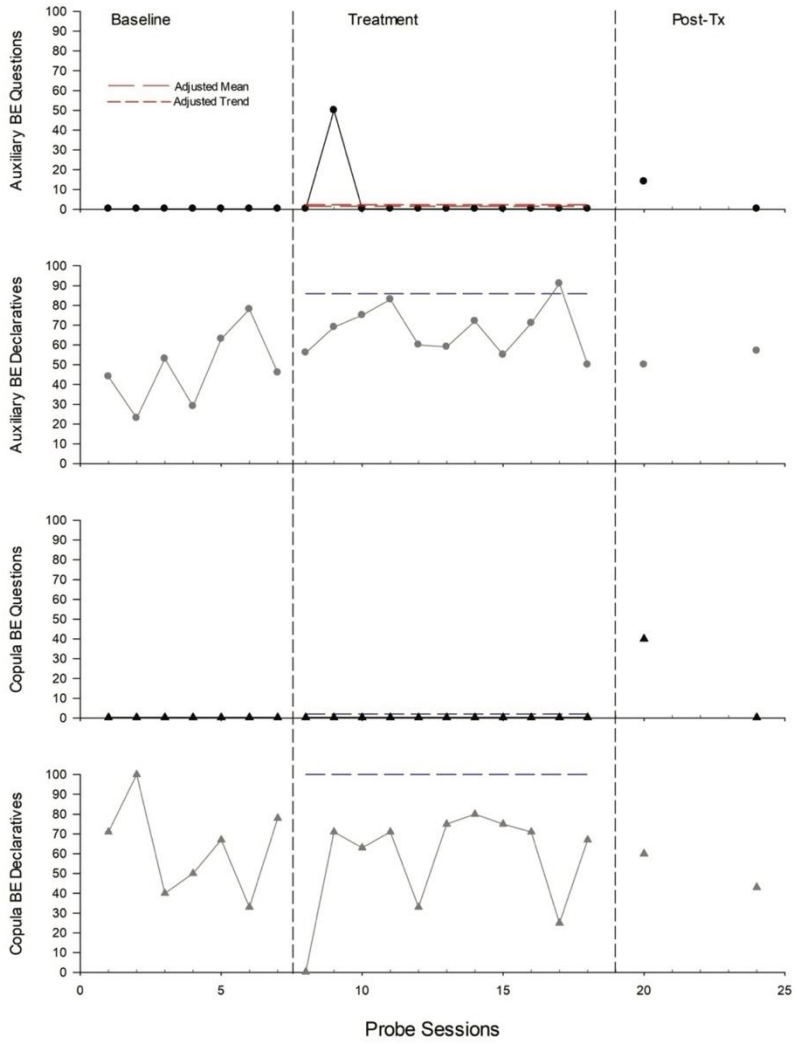
Percent Correct Auxiliary BE Questions and Untreated BE Structures on Probes for DS P1



Note. Dashed black lines indicate clinically significant change for the generalization item. Post-Tx = probes administered at 2- and 6-weeks post-treatment.

Figure 6.

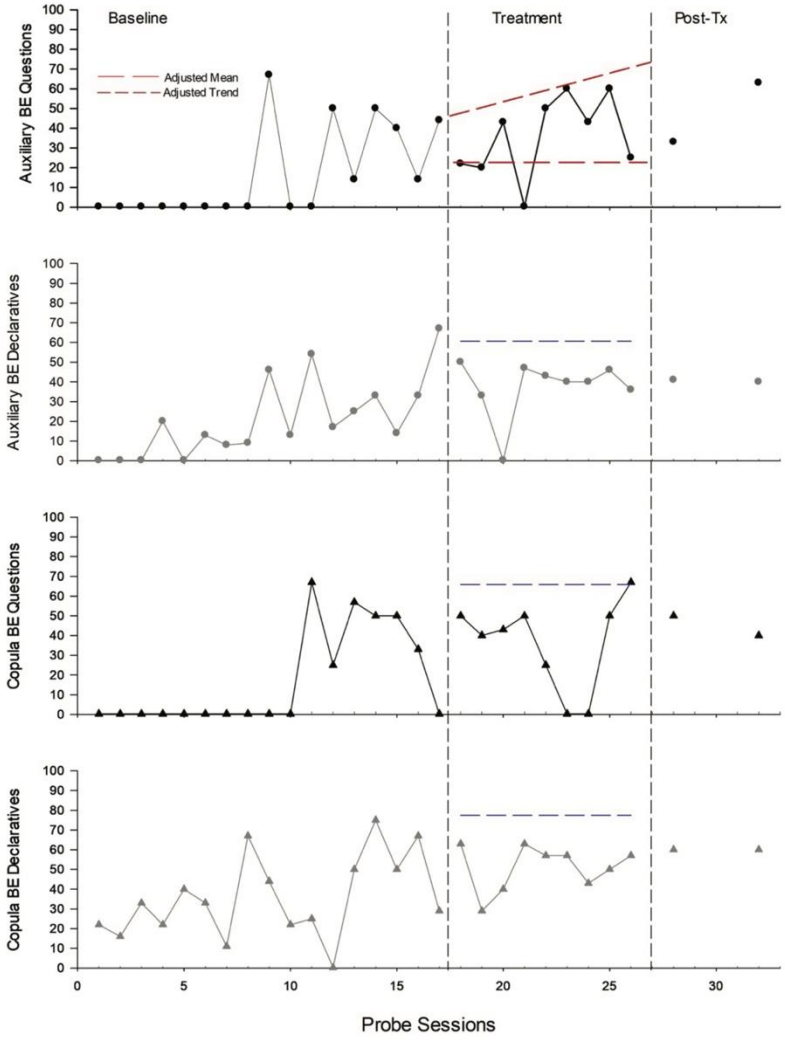
Percent Correct Auxiliary BE Questions and Untreated BE Structures on Probes for DS P2



Note. Dashed black lines indicate clinically significant change for the generalization item. Post-Tx = probes administered at 2- and 6-weeks post-treatment.

Figure 7.

Percent Correct Auxiliary BE Questions and Untreated BE Structures on Probes for DS P3



Note. Dashed black lines indicate clinically significant change for the generalization item. Post-Tx = probes administered at 2- and 6-weeks post-treatment.

Appendix A.

Report of Intervention Details According to The Single-Case Reporting Guideline in BEhavioural Interventions (SCRIBE) 2016 Checklist.

Item Number	Topic	Item Description	Our Study
TITLE and ABSTRACT			
1	Title	Identify the research as a single-case experimental design in the title	Efficacy of complexity-based target selection for treating morphosyntactic deficits in children with DLD and children with Down syndrome: A single-case experimental design
2	Abstract	Summarize the research question, population, design, methods including intervention/s (independent variable/s) and target behavior/s and any other outcome/s (dependent variable/s), results, and conclusions	Purpose: Selecting targets for morphosyntactic intervention is a critical component of treatment planning. The complexity approach suggests that, by treating a complex morphosyntactic target, improvements will occur for the treated structure and for related, simpler, structures. This study evaluated the efficacy of the complexity approach for treating morphosyntactic deficits by targeting a complex BE verb question structure for children with DLD and children with Down syndrome (DS) and observing its impact on treated and untreated BE verb structures. We also explored whether etiology impacted our participants' treatment responses. Method: Three participants with DLD and three with DS received treatment for the BE verb question structure in the context of a single-case multiple baseline design across participants. Accuracy of production for the treated structure and untreated BE verb structures was measured across baseline, treatment, and post-treatment phases. Results: Treatment of the complex BE verb question structure resulted in change on the treated structure for three participants (i.e., two with DLD and one with DS). Generalization of treatment to untreated, related BE verb structures occurred for all six participants. Outcomes indicated participants from both etiologies benefitted from treatment. Conclusions: This study provides evidence supporting the use of a complexity-based approach for selecting morphosyntactic treatment targets for children with DLD and children with DS. Additional research is needed to identify specific characteristics that may influence individual treatment responses.
INTRODUCTION			
3	Scientific background	Describe the scientific background to identify issues under analysis, current scientific knowledge, and grasp in that knowledge base	Current approaches require a large investment of time despite modest results. Complexity approach has the potential to accelerate and/or improve treatment outcomes. De Anda et al. (2020) found associations supporting the complexity approach. Etiology of language disorder has the potential to impact treatment outcomes. Children with DLD and children with DS have similar linguistic profiles and different cognitive profiles, providing an ideal opportunity to observe whether a functional relationship exists between treatment using a complexity approach and treatment outcomes. Additionally, it provides an opportunity to explore whether etiology has a clearly observable impact on treatment response across groups.

4	Aims	State the purpose/aims of the study, research questions, and if applicable, hypotheses	In this study we addressed two primary aims: (1) whether training the complex auxiliary BE question structure was causal to acquisition of that structure, (2) the extent to which training the auxiliary BE question structure resulted in generalization to untreated simpler, yet related auxiliary BE declarative, copula BE declarative, and copula BE question structures. A descriptive third aim explored whether treatment responses by children with DLD differed from children with DS. If etiology is a defining factor impacting how participants respond to treatment (Bishop et al., 2017; Catts et al., 2002; Cole & Fey, 1996), then it could be expected that clear differences in treatment responses between the small groups would be observed.
METHOD			
5	Design	Identify the design (e.g., MBD), and then describe the phases and phase sequence (a priori or data-driven) and if applicable criteria for phase change	A single-case, multiple baseline design across participants was used. Baseline, treatment, and follow-up phases (post-treatment) were incorporated into the design. A minimum baseline of 5 probes was selected a priori, based on the What Works Clearinghouse Standards 1.0 (Kratochwill et al., 2010) recommendations and our decision to use the Conservative Dual Criterion (Fisher et al., 2003) as our visual analysis strategy. Treatment phase length was response guided, but with minimum (9; to align with De Anda et al. 2020) and maximum (20) numbers of sessions determined a priori to ensure enough opportunity was provided for observation of change and to avoid unending treatment. See Supplemental Material S4 for specific detail on phase change criteria). Follow-up probes were determine a priori to occur at 2 and 6 weeks post-treatment.
6	Procedural changes	Describe any procedural changes that occurred during the course of the investigation after the start of the study	Stable baselines were not achieved for DLD P1 and DS P3 within their pre-determined minimum baseline probes, so both participants required an extension of the baseline phase.
7	Replication	Describe any planned replication	Inter-subject replication: Replication of treatment administration across 6 participants.
8	Randomization	State whether randomization was used, and if so, describe the randomization method and the elements of the study that were randomized	Within-intervention case randomization – participants randomly assigned to different tiers (i.e., P1, P2, P3, with their associated minimum baseline assignment). Story vignette order of presentation: randomized per participant. Five story vignettes were created and numbered 1-5. The order in which these were presented was randomized for each participant to avoid potential order effects. Probe: Sentences from were selected randomly from each set (e.g., 100 possible for the auxiliary BE questions) for each experimental probe.
9	Blinding	State whether blinding/masking was used, and if so, describe who was blinded/masked	Given that the study focused on treatment for school-aged children with DLD and children with DS, blinding of the primary investigator who conducted assessments and administered treatment was not possible. However, blinding with regard to the status of participant diagnoses occurred for the research assistant who participated in reliability scoring.
	PARTICIPANTS OR UNITS		

10	Selection Criteria	State the inclusion and exclusion criteria, if applicable, and the method of recruitment	Inclusionary/exclusionary criteria: DLD – age 5-8, no history of neurological damage or diagnosis of language disorder secondary to a biomedical condition (Bishop et al., 2017). DS – age 5–15, have no co-occurring autism spectrum disorder. All participants: monolingual speakers of General American English, use oral language as their primary means of communication, be receiving services from an SLP, pass a hearing and phonological screening, and present with morphosyntactic deficits, MLUw 3.0-5.0, below-criterion scores on the Test of Early Grammatical Impairment (TEGI; Rice & Wexler, 2001), and 60% or lower correct use in obligatory BE verb use on the TEGI Be/Do (Be) probe, score below 60% accuracy on at least one of the following BE structures: auxiliary BE declarative, copula BE declarative, copula BE question. Recruitment: advertisements through local school districts and Down syndrome foundations. Parents contacted the first author to indicate interest in participating in the study.
11	Participant Characteristics	For each participant, describe the demographic characteristic and clinical (or other) features relevant to the research question, such that anonymity is ensured	**Additional details provided in Table 1.** **Participants With DLD** The first participant with DLD (DLD P1) was a male aged 5;1 at entrance into the study. According to parental report, DLD P1 was evaluated at age 4 for concerns with his communication and was currently receiving 20-minute sessions weekly through the school. Participant 2 with DLD (DLD P2) was a female aged 6;11 who had been evaluated for communication concerns at age 2 and received early intervention services. She was currently receiving 30-minute sessions at school each week. Participant 3 with DLD (DLD P3) was a male aged 8;2. DLD P3 began receiving services beginning at age 2 and was currently receiving a combination of two private and two school-based sessions per week. **Participants With DS** DS participant 1 (DS P1) was a male with Down syndrome aged 12;6 at entrance into the study. He had previously received private and school-based services. He was currently receiving 60-90 minutes of therapy at the local middle school each month. Participant 2 with DS (DS P2) was a male aged 12;8. He had received services beginning in elementary school and was now receiving services twice weekly at his local middle school. Participant 3 with DS (DS P3) was a female aged 7;3. She began receiving services at age 2. At the time of the study, she was receiving SLP services at school 20 minutes weekly.
	CONTEXT		
12	Setting	Describe characteristics of the setting and location where the study was conducted	All sessions occurred in the homes of the participants. Eligibility assessments: In-person interactions included obtaining parental consent and participant assent, the initial interview, a hearing screening, cognitive testing (i.e., WNV), and a narrative language sample collection. The GFTA–3, PPVT–4, and the TEGI assessments were administered remotely over Zoom. Baseline, treatment, and post-treatment probe data for treated and untreated items were all collected remotely. All treatment sessions were delivered via remote means (i.e., Zoom). Post-treatment assessments were collected via remote means.

	APPROVALS		
13	Ethics	State whether ethics approval was obtained and indicate if and how informed consent and/or assent were obtained	Approval for all aspects of the study was provided by the institutional review board at the University of Utah. Parents and the participants were provided information regarding the study via written and verbal means. Parents provided written informed consent and participants provided assent prior to beginning eligibility assessments.
	MEASURES and MATERIALS		
14	Measures	Operationally define all target behaviors and outcome measures, describe reliability and validity, state how they were selected, and how and when they were measured	**Independent variable:** treatment of auxiliary BE questions **Dependent variables:** The primary dependent variable was the production of auxiliary BE questions on experimental probes. The secondary dependent variables were the production of copula BE questions and auxiliary and copula BE declaratives in experimental probes. All probes consisted of stimuli using BE present tense verbs (e.g., *is, are*). Variable measurement defined as percent accurate production of auxiliary BE questions, auxiliary BE declaratives, copula BE questions, copula BE declaratives. Outcomes were calculated per probe as the total correct attempts at each structure divided by the total number of correct and incorrect attempts at each structure. The denominator was variable to accommodate the inherent flexibility required with elicited probe conditions. This practice was in alignment with De Anda et al. (2020), which we were trying to replicate. As such, the percent correct reported for each probe was calculated with a denominator representing both incorrect and correct attempts for each BE verb structure, and a numerator representing only the number of correct attempts. Although we presented elicited production opportunities for each structure with consistency, the nature of the task allowed the participants flexibility with the type of structure they used in their responses. Therefore, the denominator for percent correct varied. For purposes of transparency with our data, we have created bar graphs that show the number of correct and incorrect attempts for each percent correct calculation per probe, per participant, and included them as Supplemental Material S5. **Reliabiility:** Point-to-point scoring reliability was completed for 20% of the administered probes per phase and per participant. A research assistant, blinded to participant status, scored the probes using session video recordings. Cohen's κ was calculated to determine the extent to which the first and second scorers agreed on probe items. Results indicated a substantial agreement (Landis & Koch, 1977) between scorer judgments, $\kappa = .715$ (95% CI, .670 to .760), $p < .001$. **Validity:** Probes were experimental probes developed following protocols utilized by De Anda et al. (2020). **Frequency of measurement:** Probes were administered at each session during baseline and treatment phases. Probes were administered prior to beginning any treatment session to mitigate possible immediate treatment effects. Probes were also administered during the post-treatment phases at 2- and 6-weeks as follow-up to determine retention of trained skills and generalized skills.
15	Equipment	Clearly describe any equipment and/or materials used to measure	**Materials for experimental probes:** Probes were developed to elicit treated and untreated structures using picture stimuli and verbal prompts. 100 lexical verbs were used to develop auxiliary BE question and declarative probes and 60 modifiers were used to develop copula BE question and declarative

		target behaviors and other outcomes or deliver the interventions	probes. See manuscript (i.e., Methods; Experimental Stimuli) for complete description of these materials. **Materials for intervention:** Intervention protocols utilized five story vignette scripts, three stuffed animals, and props to create the storyline associated with the story vignette script (e.g., toothbrush, toy food, hats, shoes, balls). Participants utilized a computer and speakers to sign on to their individual Zoom session.
	INTERVENTIONS		
16	Intervention	Describe the intervention and control condition in each phase, including how and when they were actually administered, with as much detail as possible to facilitate attempts at replication	Intervention protocol (i.e., intervention script and cueing hierarchy tracking document) provided as Appendix A. Intervention sessions were conducted via Zoom. Story vignettes provided a scripted protocol that was followed for each intervention session. Fifteen different verbs were used twice in each story vignette—once in a singular context and once in a plural context—for a total of 30 teaching episodes. In other words, participants received a dose of 30 teaching episodes per 15-to-30-minute session (Warren et al., 2007). Total session length was variable based on the amount of cueing that was required. Parents helped participants sign on to their private Zoom session from a quiet place in their homes. Story vignettes incorporated two mice and a dog, who "completed" tasks relatable to the participant (e.g., getting ready for school). The participant's role was to ask the puppet questions about the mice and dog. The interventionist prompted for each teaching episode according to a scripted protocol (Appendix A). For example, the interventionist would show the dog brushing his teeth and say *"I wonder if the dog is brushing. Ask if the dog is brushing."* A graduated cueing hierarchy (Appendix A) was used to ensure participants' success with teaching episodes and included three levels: (1) a repetition of the prompt, (2) a model of the targeted response (e.g., "*I'll ask the puppet: is the mouse eating? Now you do it*"), and (3) an elicited imitation (e.g., "*Say, is the mouse eating.*"). The following incorrect responses initiated the cueing hierarchy: (a) no response, (b) no attempt of an auxiliary BE question, (c) use of the wrong auxiliary verb form (e.g., *is* for *are*), or (d) omission of the auxiliary BE form (e.g., *The dog eating?*). Control conditions included: same mice and dog as characters for each story vignette; two sessions administered weekly; story vignettes were repeated up to four times each in a pre-determined, randomized sequence per participant; cueing hierarchy was controlled and initiated for each teaching episode as needed.
17	Procedural fidelity	Describe how procedural fidelity was evaluated in each phase	Procedural fidelity was evaluated for 20% of each experimental probe per participant and per phase by an independent second observer. Fidelity of the interventionist (first author) to the treatment protocol was evaluated for 10% of the treatment sessions per participant by a second observer. The number of correctly provided treatment components divided by the total number expected resulted in a treatment fidelity estimate of 100%.
	ANALYSIS		
18	Analyses	Describe and justify all methods used to analyze data	For a complete description of the analytic plan, see the manuscript: Data Analysis Plan. Summary description: visual analysis used the Conservative Dual Criterion method (CDC; Fisher et al.,2003) for the primary dependent variable. Tau-U was used to evaluate the magnitude of change for participants through statistical analysis of the primary dependent variable. Secondary dependent

			variables were evaluated arithmetically through the calculation of clinically significant change above baseline levels (i.e., 2 standard deviations above the mean during baseline).
RESULTS			
19	Sequence completed	For each participant, report the sequence actually completed, including the number of trials for each session for each case. For participants who did not complete, state when they stopped and the reasons	All participants completed all phases of the study. Participants who did not receive 20 treatment sessions met a priori criteria for moving to the post-treatment phase early. See the manuscript description of these criteria in Method; Treatment Phase, and Supplemental Material S4 for the flowchart. DLD P1: Baseline – 10 probe sessions; Treatment – 20 treatment sessions; Post-treatment – 2 probe sessions DLD P2: Baseline – 7 probe sessions; Treatment – 20 treatment sessions; Post-treatment – 2 probe sessions DLD P3: Baseline – 9 probe sessions; Treatment – 20 treatment sessions; Post-treatment – 2 probe sessions DS P1: Baseline – 5 probe sessions; Treatment – 20 treatment sessions; Post-treatment – 2 probe sessions DS P2: Baseline – 7 probe sessions; Treatment – 11 treatment sessions; Post-treatment – 2 probe sessions DS P3: Baseline – 17 probe sessions; Treatment – 9 treatment sessions; Post-treatment – 2 probe sessions
20	Outcomes and estimation	For each participant, report results, including raw data, for each target behavior and other outcomes	Our results are provided for each probe as percent accurate productions, to standardize participant responses. Additional transparency was necessary due to the changing denominators on the probes. Therefore, graphs showing the frequency of correct and total attempts by each participant per probe and per phase are also provided as Supplemental Material S5.
21	Adverse events	State whether or not any adverse events occurred for any participant and the phase in which they occurred	No adverse events occurred for any participant during the study.
DISCUSSION			
22	Interpretation	Summarize findings and interpret the results in the context of current evidence	Results support the efficacy of a complexity-based approach to treatment of morphosyntax for BE verb structures. See Discussion section for more detail.
23	Limitations	Discuss limitations, addressing sources of potential bias and imprecision	See Limitations subsection of Discussion in manuscript.
24	Applicability	Discuss applicability and implications of the study findings	Using complexity-based treatment targets has potential to support linguistic learning in participants with DLD and with DS. See Discussion section for more detail.

DOCUMENTATION			
25	Protocol	If available, state where a study protocol can be accessed	Intervention protocol (i.e., intervention script and cueing hierarchy tracking document) provided as Appendix B.
26	Funding	Identify sources of funding and other support; describe the role of funders	This project was unfunded.

897
898

Appendix B.

Example Story Vignette Treatment Script and Cueing Hierarchy Tracker.

INTRO:
We're going to play a game. In this game we have two mice, a dog, a cat and puppet. We can't talk to the animals because they speak a silly language that we don't know. But the puppet speaks the silly animal language. So, if we want to know things about the animals, we have to ask the puppet.

Let me show you what I mean. I'll ask the puppet a question:
Examiner (E): Can you dance?
Puppet (P): Yes! (puppet dances)

TRAINING:
Now you try it. Ask the puppet a question. Ask the puppet if he can talk.
Target (T): Can you talk?
See? We can talk to the puppet and he can understand us. So, if we want to know things about the mice and the dog we can ask the puppet. Are you ready?

Tx Item		Script	Level of Prompting			
			None	Repetition	Model	Imitation
1		*(mice stand in front of some food)*				
	E	I wonder if the mice are making breakfast. Ask if the mice are making breakfast.				
	T	***Are the mice making breakfast?***				
	P	Yes.				
2		(*dog moves to the food)*				
	E	And I wonder if the dog is making breakfast. Ask if the dog is making breakfast.				
	T	***Is the dog making breakfast?***				
	P	Yes.				
3		*(mice move to the pan)*				
	E	I wonder if the mice are cooking their breakfast. Ask if the mice are cooking.				
	T	***Are the mice cooking?***				
	P	Yes.				
4		*(dog moves to the pan)*				
	E	The dog wants to cook too. Ask if the dog is cooking.				
	T	***Is the dog cooking?***				
	P	Yes.				
5		*(mice make smelling movements)*				
	E	Look at the mice. I wonder if the mice are smelling the food. Ask if the mice are smelling.				
	T	***Are the mice smelling?***				
	P	Yes!				
6		*(dog makes smelling movements)*				
	E	And I wonder about the dog. Ask if the dog is smelling the food.				
	T	***Is the dog smelling?***				
	P	Yes! Mmm the food smells good.				

Tx Item		Script	Level of Prompting			
			None	Repetition	Model	Imitation
7	E	All done cooking! Now it's time to taste the food. *(mice make tasting movements)*				
	E	I wonder if the mice are tasting the food. Ask if the mice are tasting.				
	T	***Are the mice tasting?***				
	P	Yes! Ouch it's hot!				
8		*(dog makes tasting movements)*				
	E	And I wonder if the dog is tasting his food. Ask if the dog is tasting.				
	T	***Is the dog tasting?***				
	P	Yes. His food is hot too!				
9	E	I think they need to cool down their food. *(mice blow on food)*				
	E	I wonder if the mice are blowing on their food. Ask if the mice are blowing on the food.				
	T	***Are the mice blowing on the food?***				
	P	Yes.				
10		*(dog blows on his food)*				
	E	And I wonder if the dog is blowing on his food. Ask if the dog is blowing on the food.				
	T	***Is the dog blowing on the food?***				
	P	Yes.				
11	E	Oh, I think the food has cooled down. How do they eat? Do they eat like people? No. I think they lick their food first. *(mice make licking sounds)*				
	E	I wonder if the mice are licking their food? Ask if the mice are licking their food.				
	T	***Are the mice licking their food?***				
	P	Yes.				
12		*(dog makes licking sounds)*				
	E	And I wonder if the dog is licking the food. Ask if the dog is licking the food.				
	T	***Is the dog licking the food?***				
	P	Yes.				
13	E	What do they do after they lick their food? I think they bite it. *(biting sounds)*				
	E	I wonder if the mice are biting their food. Ask if the mice are biting their food.				
	T	***Are the mice biting the food?***				
	P	Yes.				
14	E	Maybe the dog does the same thing. I wonder if the dog is biting his food. Ask if the dog is biting his food.				
	T	***Is the dog biting his food?***				
	P	Yes.				
15	E	Okay. Look! The cat hasn't eaten. (*food moves to the cat)*				
	E	I wonder if the cat is eating the food. Ask if the cat is eating.				
	T	***Is the cat eating?***				
	P	No.				
16	E					

Tx Item		Script	Level of Prompting			
			None	Repetition	Model	Imitation
		Oh no! Maybe the mice can show him that it tastes good. I wonder if the mice are eating his food. Ask if the mice are eating his food.				
	T	***Are the mice eating his food?***				
	P	Yes.				
17	E	Maybe the cat is thirsty. *(pull out orange juice*)				
	E	Here's some orange juice. *(cat shakes orange juice)*				
	E	I wonder if the cat is shaking the juice. Ask if the cat is shaking the juice.				
	T	***Is the cat shaking the juice?***				
	P	Yes.				
18	E	The mice want to shake the juice too. *(mice shake juice)*				
	E	I wonder if the mice are shaking the juice. Ask if the mice are shaking the juice.				
	T	***Are the mice shaking the juice?***				
	P	Yes.				
19	E	Alright, the juice is ready. *(mice pour juice)*				
	E	I wonder if the mice are pouring the juice into their cups. Ask if the mice are pouring the juice.				
	T	***Are the mice pouring the juice?***				
	P	Yes.				
20		*(cat pours juice*)				
	E	And I wonder if the cat is pouring the juice. Ask if the cat is pouring the juice.				
	T	***Is the cat pouring the juice?***				
	P	Yes.				
21		*(mice drop the juice)*				
	E	Oh no! I wonder if the mice are spilling the juice. Ask if the mice are spilling the juice.				
	T	***Are the mice spilling the juice?***				
	P	Yes.				
22		*(dog picks up juice and drops it too)*				
	E	And I wonder if the dog is spilling the juice. Ask if the dog is spilling.				
	T	***Is the dog spilling the juice?***				
	P	Yes.				
23	E	Good! Now it's time to finish breakfast. I wonder if the mice are finishing breakfast. Ask if the mice are finishing.				
	T	***Are the mice finishing?***				
	P	Yes.				
24	E	And I wonder if the dog is finishing his breakfast. Ask if the dog is finishing.				
	T	***Is the dog finishing?***				
	P	Yes.				
25	E	Whew, what a mess! Now it's time to clean up. I wonder if the mice are cleaning the mess. Ask if the mice are cleaning.				
	T	***Are the mice cleaning?***				

Tx Item		Script	Level of Prompting			
			None	Repetition	Model	Imitation
	P	Yes.				
26	E	Everyone needs to clean. I wonder if the dog is cleaning. Ask if the dog is cleaning.				
	T	***Is the dog cleaning?***				
	P	Yes.				
27	E	What kind of cleaning do they do? *(animals are wiping the table)*				
	E	I wonder if the mice are wiping the table. Ask if the mice are wiping.				
	T	***Are the mice wiping?***				
	P	Yes.				
28	E	And I wonder if the cat is wiping the table. Ask if the cat is wiping.				
	T	***Is the cat wiping?***				
	P	Yes.				
29	E	Now it's time to clean the dishes. *(animals move to the dishes)*				
	E	I wonder if the mice are washing dishes. Ask if the mice are washing.				
	T	***Are the mice washing?***				
	P	Yes.				
30	E	And I wonder if the dog is washing dishes. Ask if the dog is washing.				
	T	***Is the dog washing?***				
	P	Yes. Breakfast is all cleaned up now. Great job everyone. The end.				

Supplemental Material S1

Sample experimental probe for auxiliary BE questions

Supplemental Material S2

Sample experimental probe for untreated structures

Supplemental Material S3

Frequency of correct and incorrect attempts for each BE verb probe and per participant

Supplemental Material S4

Flowchart for determining continuation or completion of treatment

Supplemental Material S1. Sample Experimental Probe for Auxiliary BE Questions.

INTRO:

Here is my friendly puppet. We are going to look at some pictures with him. He is a special puppet. We can understand the puppet. So, we can talk to him and ask him questions about the pictures. Let me show you.

Training:

E: *Look at this picture. I wonder if the cow can moo. Ask the puppet.*

Target: *"(Puppet) can the cow moo?"*

If the participant does not provide a question, provide the target and ask them to imitate it. Then re-administer the prompt.

E: *Great. Let's look at some more pictures.*

Prompts: *"Yes. Can you say it a different way?"*

Prompts are only provided if there is: (a) no response or, (b) no attempt to use a BE verb (e.g., *"She likes the cookie"*). No other feedback or cueing is allowed during this portion of the probe.

Prompt	Target	Response	Correct?
I wonder if the boys are smiling. Ask the puppet.	Are the boys smiling?		
I wonder if the girls are washing their hands. Ask the puppet.	Are the girls washing their hands?		
I wonder if the boys are talking. Ask the puppet.	Are the boys talking?		
I wonder if the boys are pushing the swing. Ask the puppet.	Are the boys pushing the swing?		
I wonder if the girl is thinking. Ask the puppet.	Is the girl thinking?		
I wonder if the rabbit wiping the window. Ask the puppet.	Is the rabbit wiping the window?		
I wonder if the dog is looking at the plane. Ask the puppet.	Is the dog looking at the plane?		
I wonder if the boy is sweeping the floor. Ask the puppet.	Is the boy sweeping the floor?		
I wonder if the boys are reading. Ask the puppet.	Are the boys reading?		
I wonder if the girl is playing. Ask the puppet.	Is the girl playing?		
		Total Correct	

Supplemental Material S2. Sample Experimental Probe for Untreated Structures.

INTRO:

We are going to look at some more pictures.

Prompts: *"Yes. Can you say it a different way?"*

Prompts are only provided if there is: (a) no response or, (b) no attempt to use a BE verb (e.g., *"She likes the cookie"*). No other feedback or cueing is allowed during this portion of the probe.

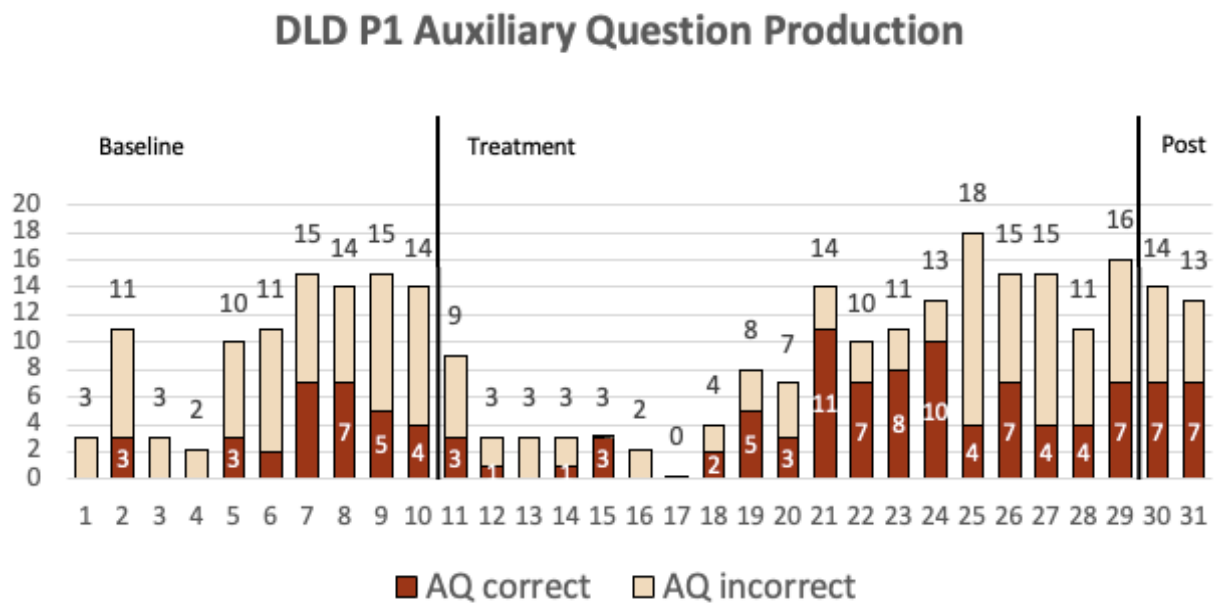
Prompt		Target	Response	Correct?
Copula BE questions				
P26	I wonder if the rocks are hard. Ask the puppet.	Are the rocks hard?		
S12	I wonder if the toy is broken. Ask the puppet.	Is the toy broken?		
S09	I wonder if the dog is cold. Ask the puppet.	Is the dog cold?		
P11	I wonder if the boys are hungry. Ask the puppet.	Are the boys hungry?		
P02	I wonder if the girls are wet. Ask the puppet.	Are the girls wet?		
S16	I wonder if the ball is orange. Ask the puppet.	Is the ball orange?		
			Total Correct	

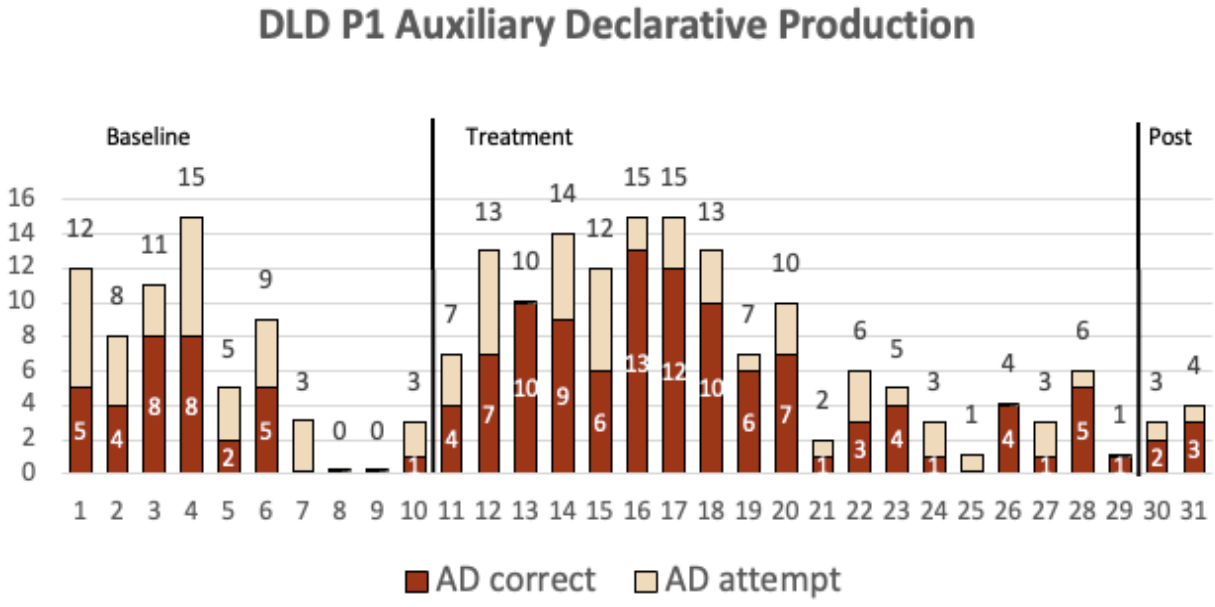
Prompt		Target	Response	Correct?
Copula Be sentences				
P27	Tell me about the stars.	The stars are yellow.		
S18	Tell me about the girl.	The girl is happy.		
P02	Tell me about the girls.	The girls are sticky.		
P24	Tell me about the dogs.	The dogs are asleep.		
S30	Tell me about the shoe.	The shoe is clean.		
S12	Tell me about the dog.	The dog is thirsty.		
			Total Correct	
Auxiliary BE sentences				
P08	Tell me what's happening with the cats.	The cats are tying bows.		
S05	Tell me what's happening with the cat.	The cat is brushing her fur.		
S13	Tell me what's happening with the rabbit.	The rabbit is watching tv.		
P23	Tell me what's happening with the rabbits.	The rabbits are clapping.		
S13	Tell me what's happening with the rabbit.	The rabbit is biting the cheese.		
P09	Tell me what's happening with the rabbits.	The rabbits are skating.		
			Total Correct	

Supplemental Material S3. Frequency of Correct and Incorrect Attempts for Each BE Verb Probe and Per Participant.

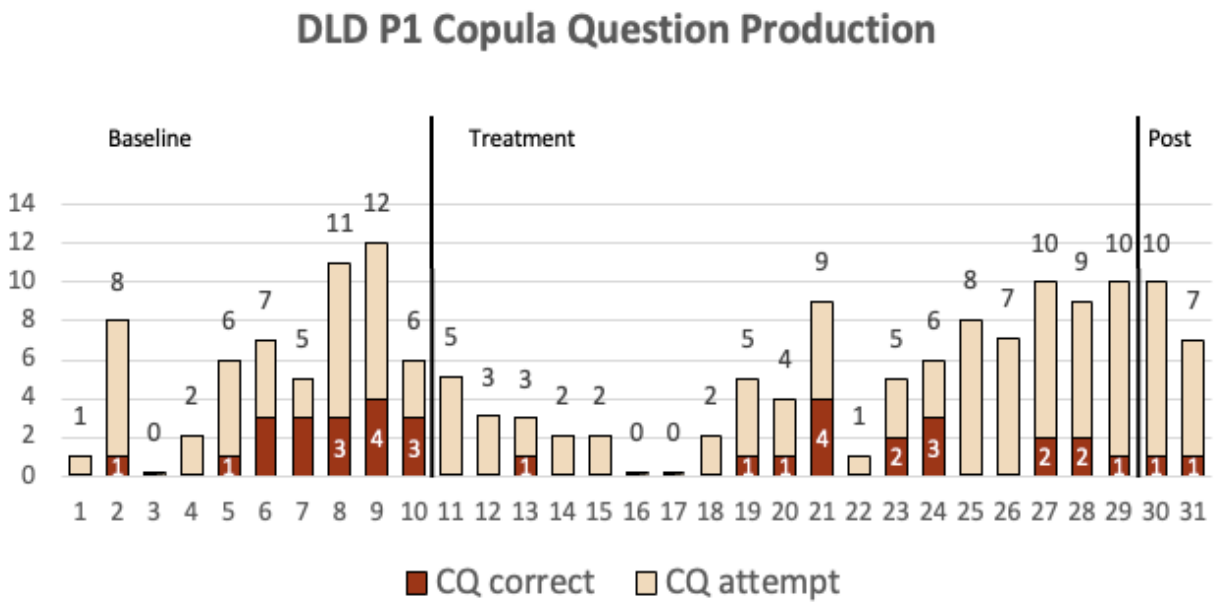
AQ = Auxiliary Questions. AD = Auxiliary Declaratives. CQ = Copula Questions. CD = Copula Declaratives.

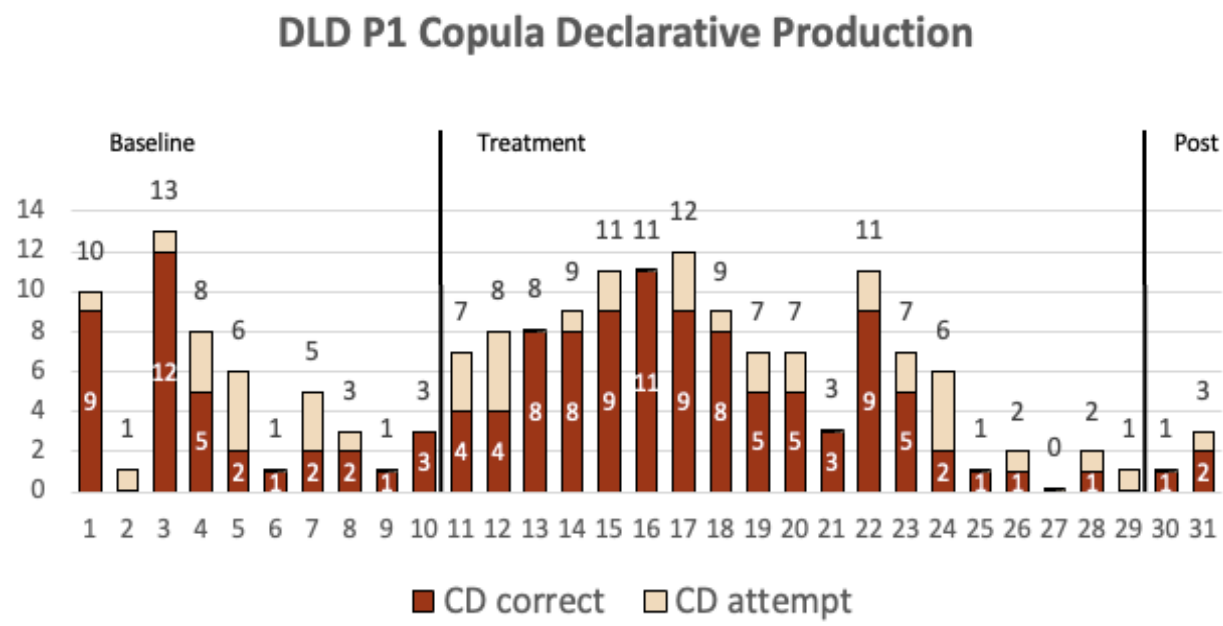
DLD P1:



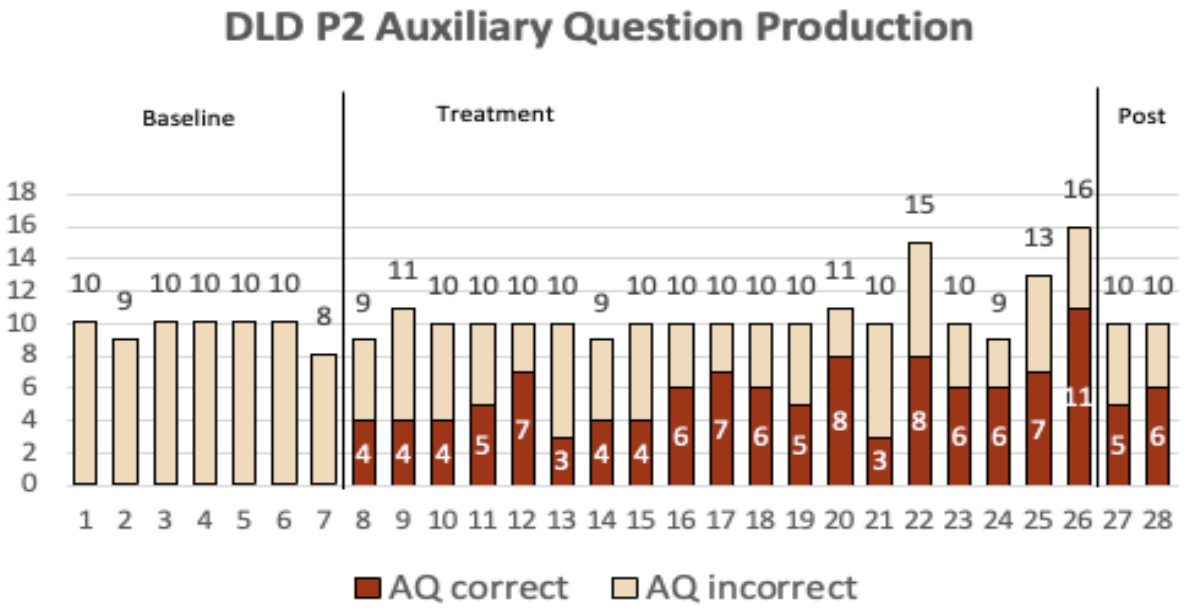


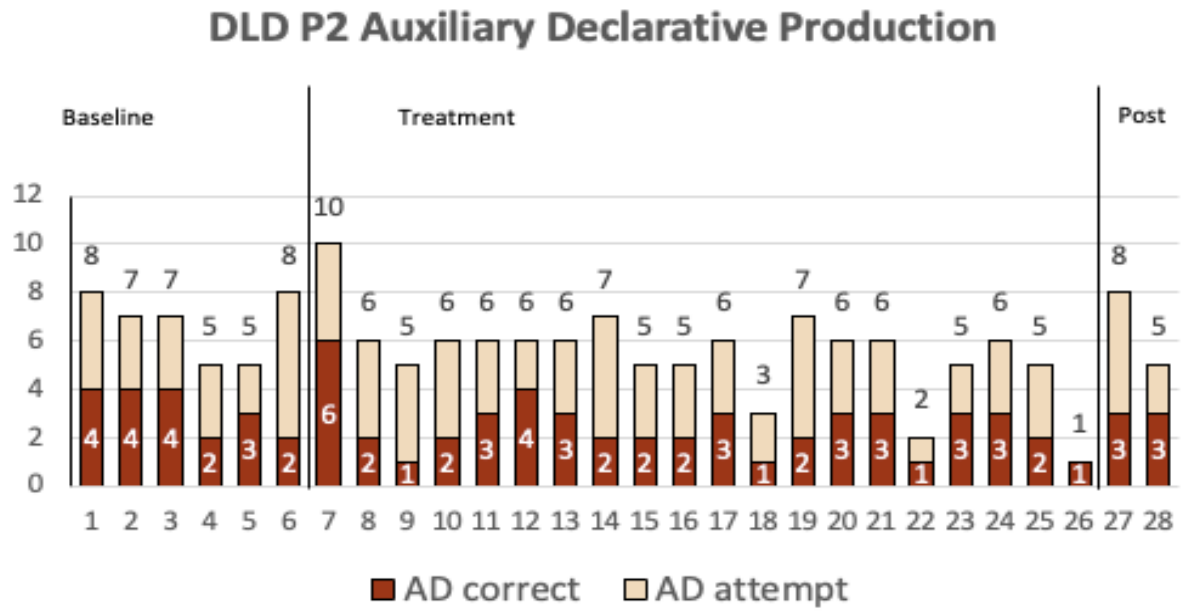
991

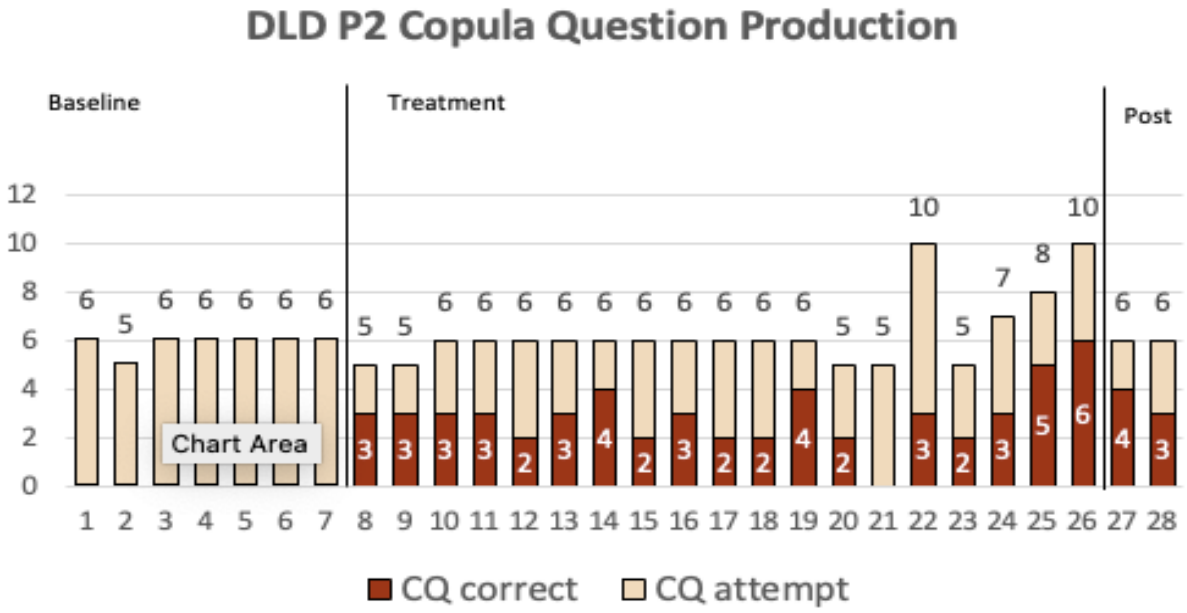

DLD P1 Copula Question Production
Baseline
Treatment
Post
CQ correct
CQ attempt

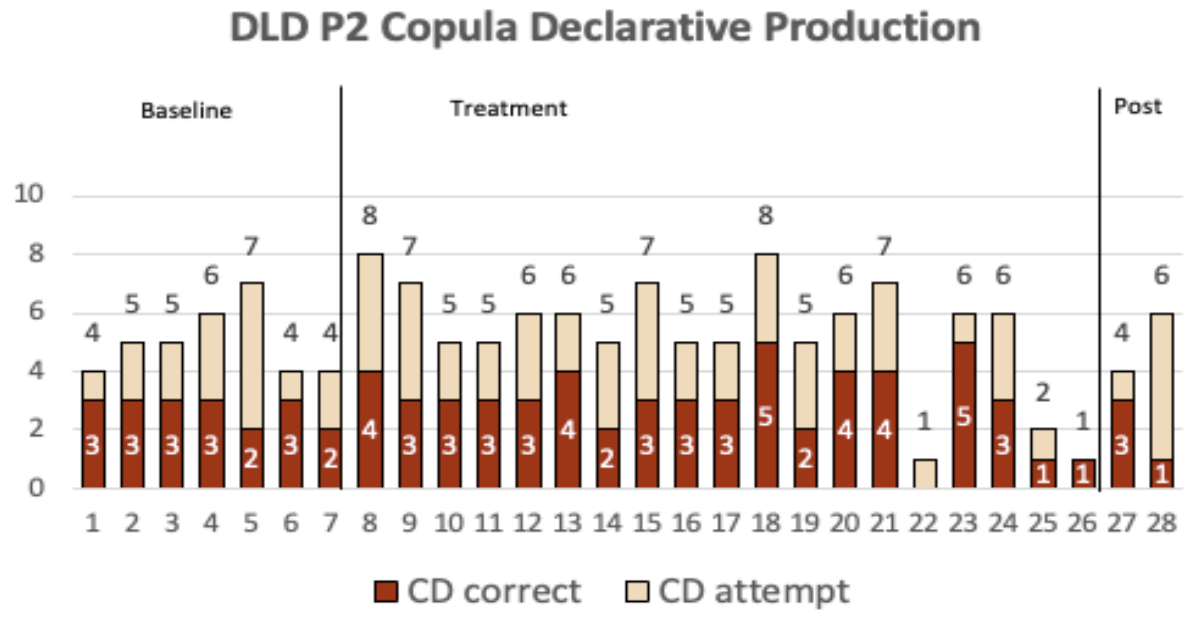

DLD P1 Copula Declarative Production
Baseline
Treatment
Post
CD correct
CD attempt

DLD P2:

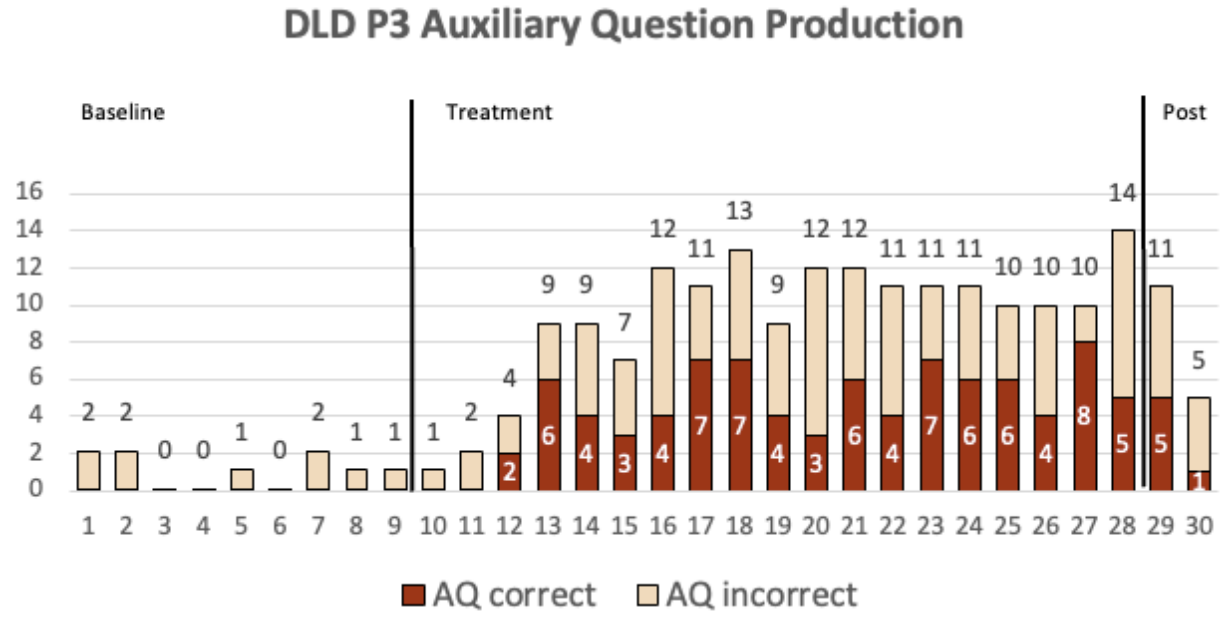

DLD P2 Auxiliary Question Production
Baseline
Treatment
Post
AQ correct
AQ incorrect

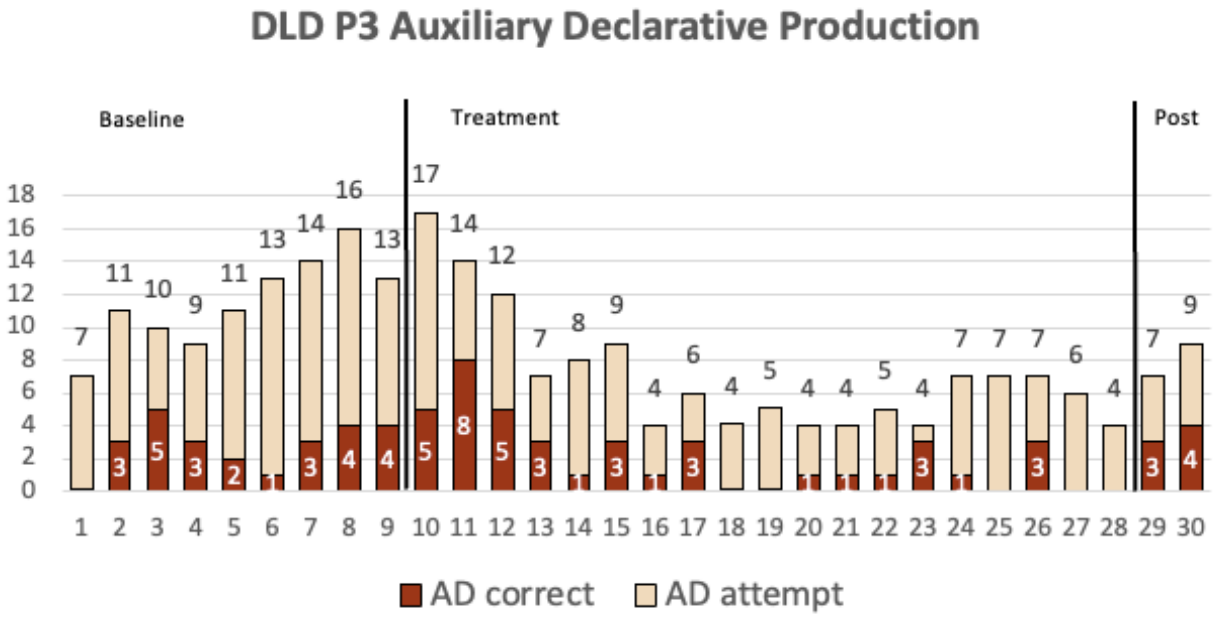

DLD P2 Auxiliary Declarative Production
Baseline
Treatment
Post
AD correct
AD attempt


DLD P2 Copula Question Production
Baseline
Treatment
Post
Chart Area
CQ correct
CQ attempt

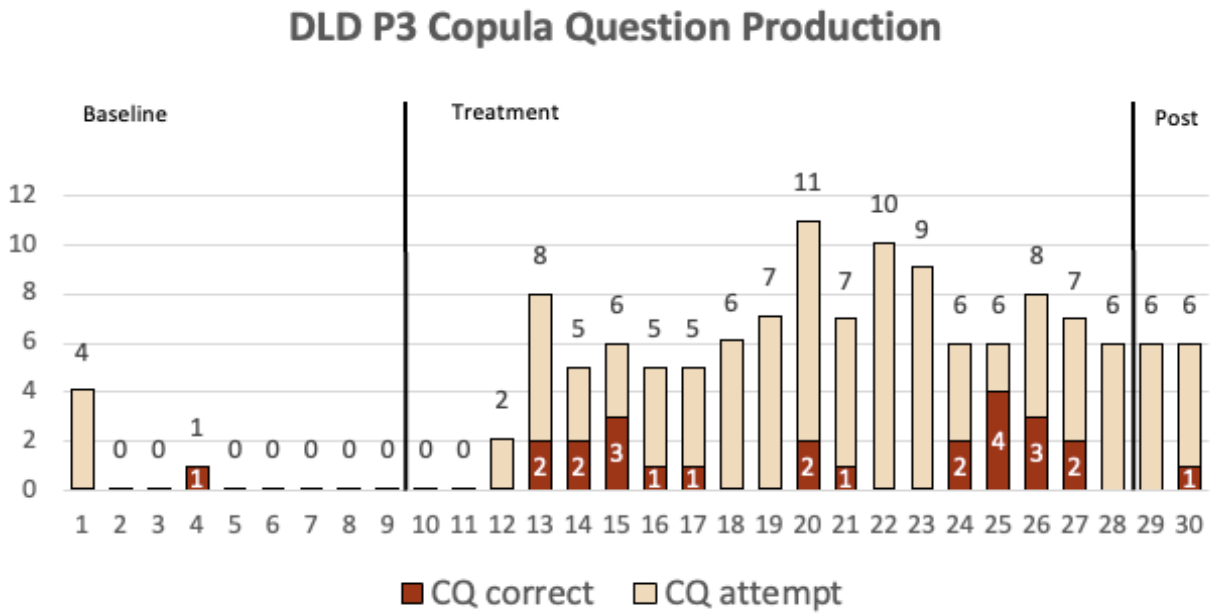

DLD P2 Copula Declarative Production
Baseline
Treatment
Post
CD correct
CD attempt

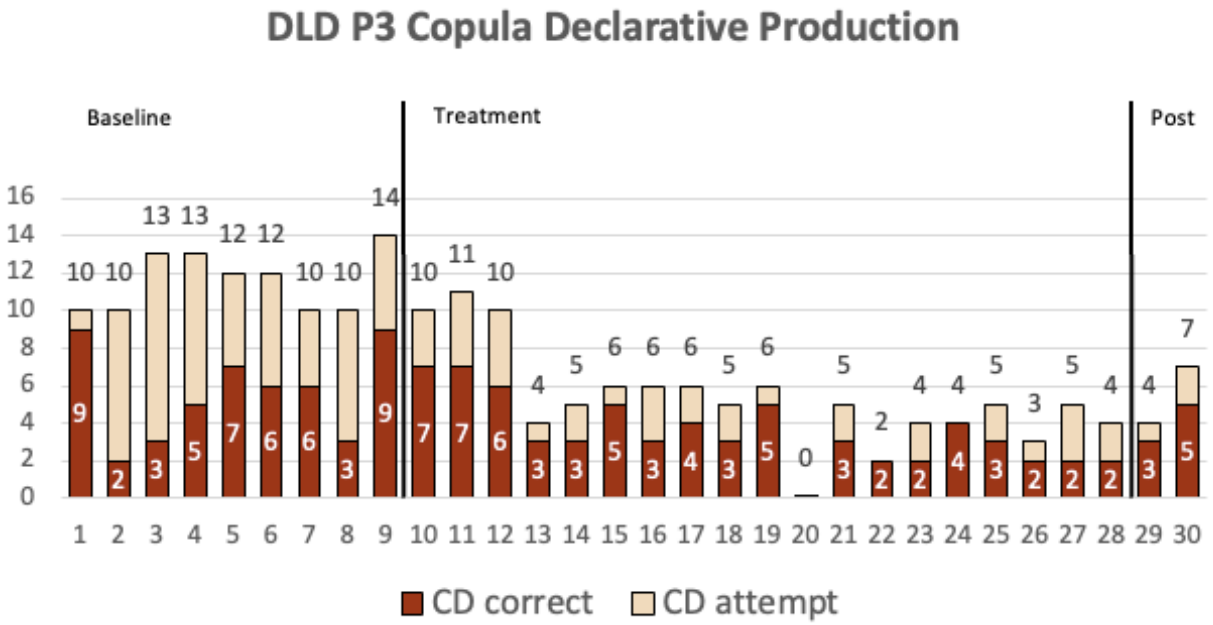
DLD P3:


DLD P3 Auxiliary Question Production
Baseline
Treatment
Post
AQ correct
AQ incorrect

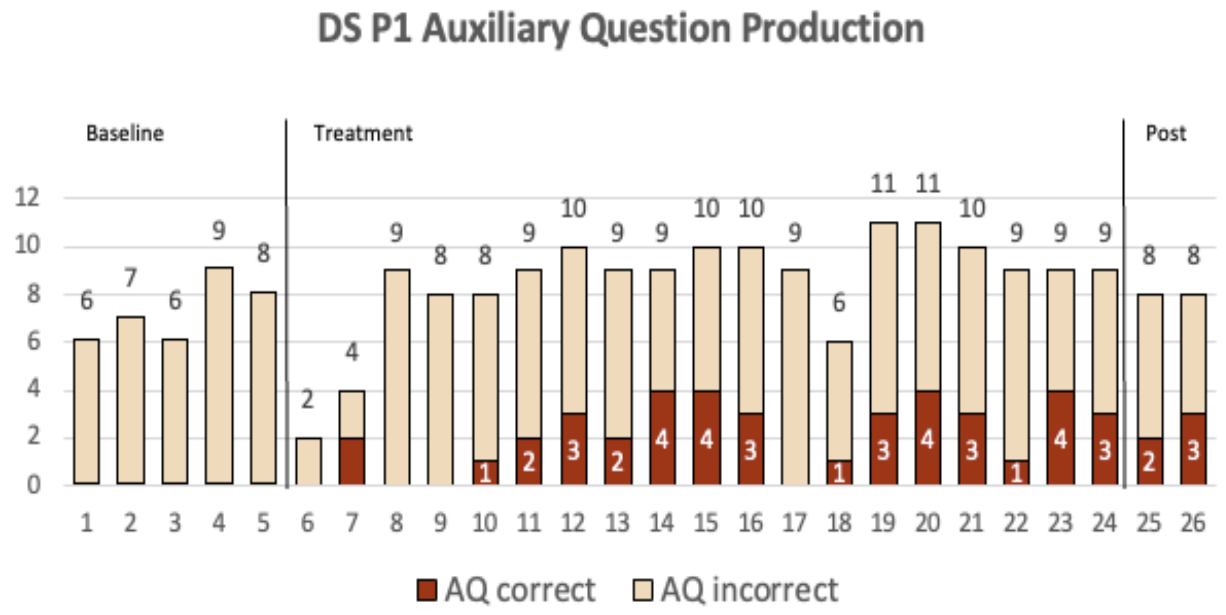

DLD P3 Auxiliary Declarative Production
Baseline
Treatment
Post
AD correct
AD attempt

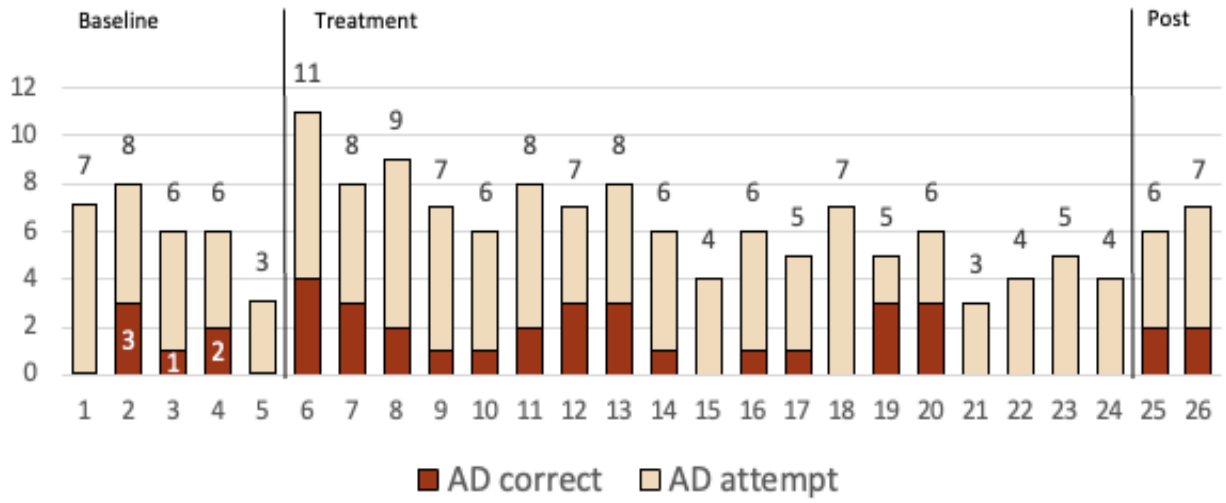
1005


DLD P3 Copula Question Production
Baseline
Treatment
Post
CQ correct
CQ attempt

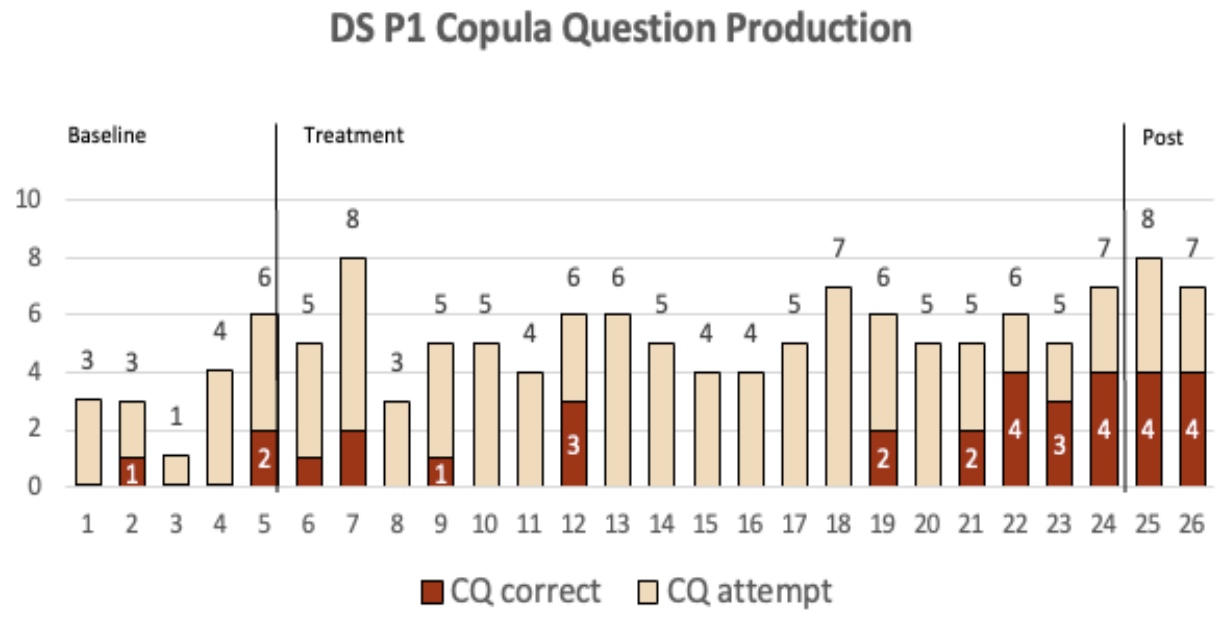

DLD P3 Copula Declarative Production
Baseline
Treatment
Post
CD correct
CD attempt

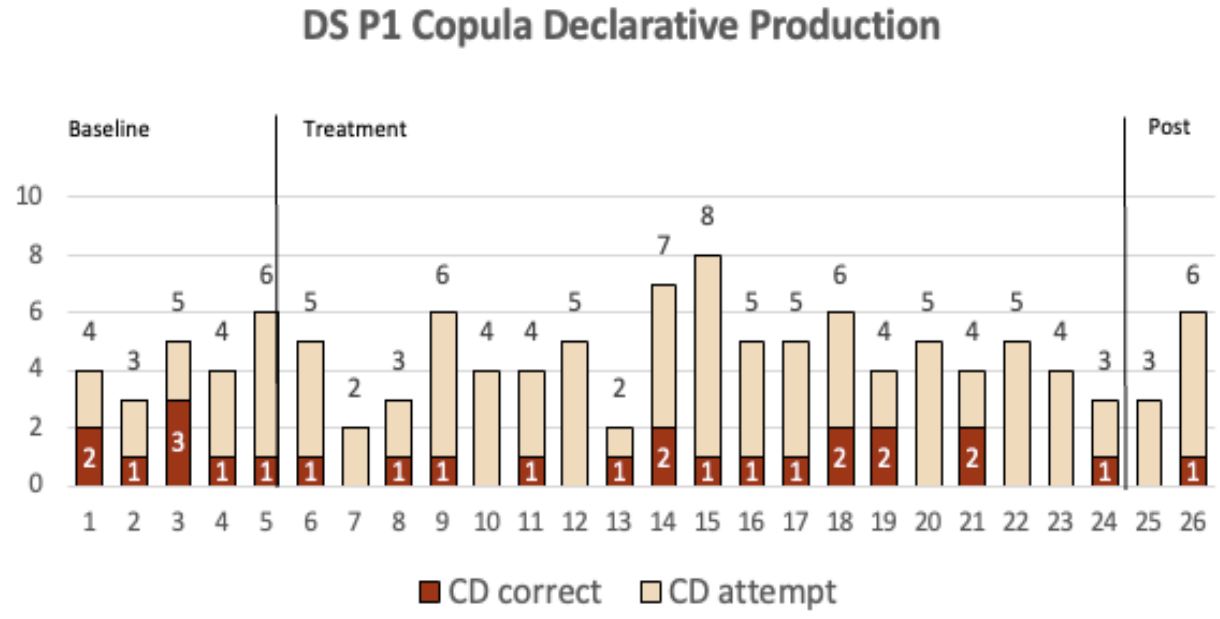
DS P1:


DS P1 Auxiliary Question Production
Baseline
Treatment
Post
AQ correct
AQ incorrect

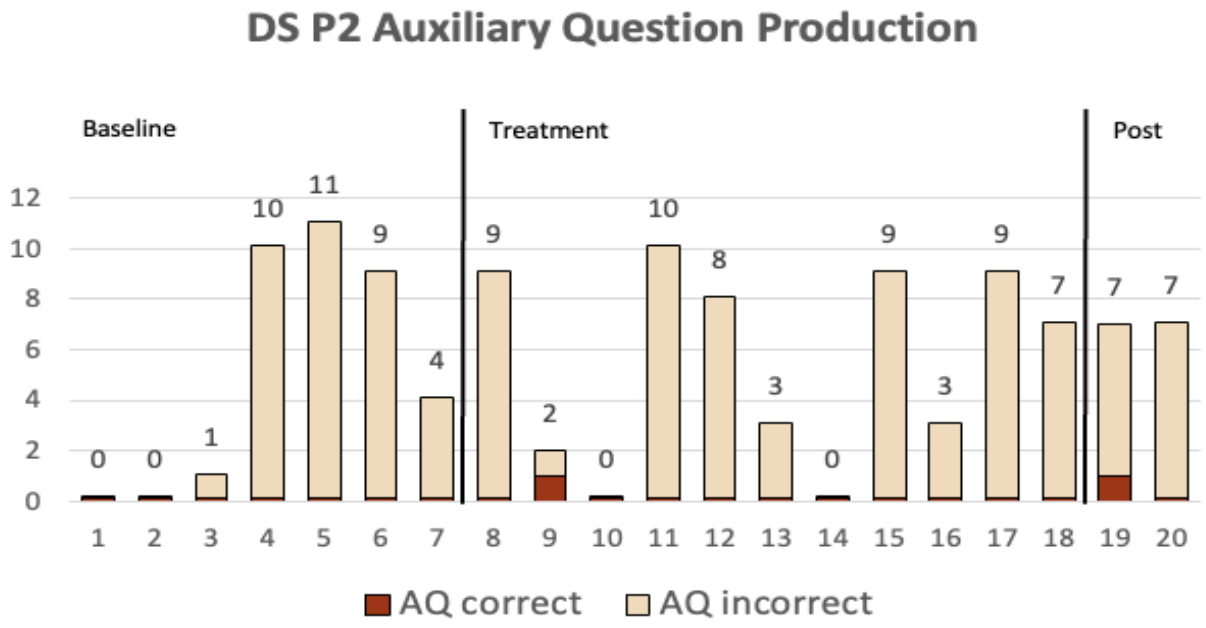

DS P1 Auxiliary Declarative Production
Baseline
Treatment
Post
AD correct
AD attempt

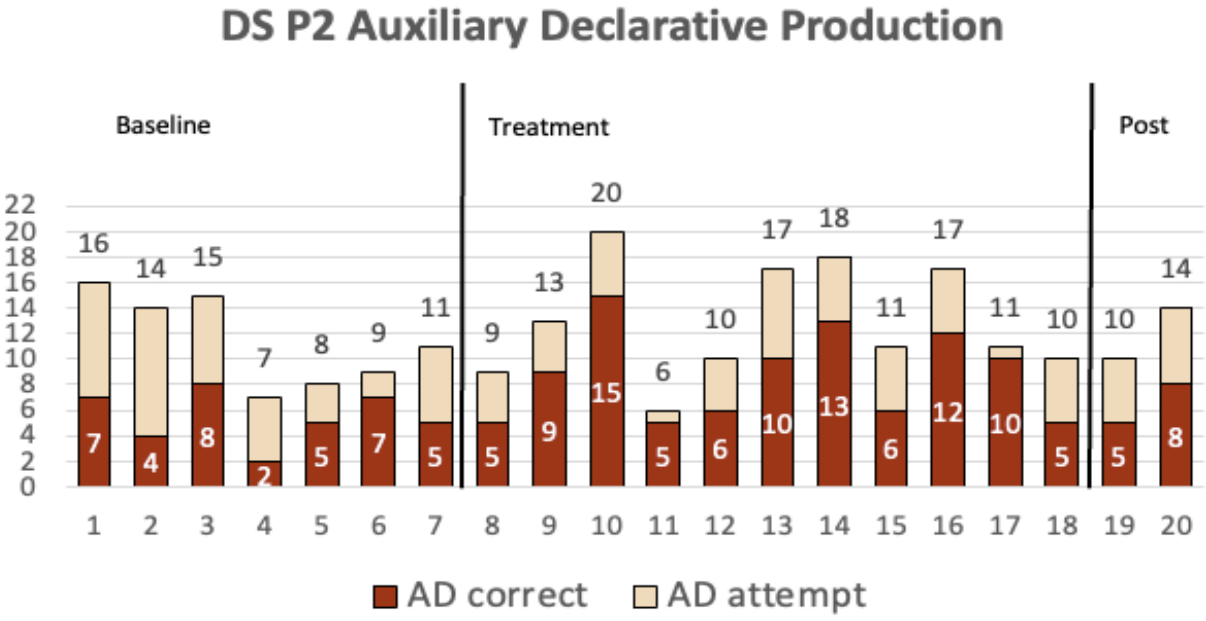
1008
1009


DS P1 Copula Question Production
Baseline
Treatment
Post
CQ correct
CQ attempt

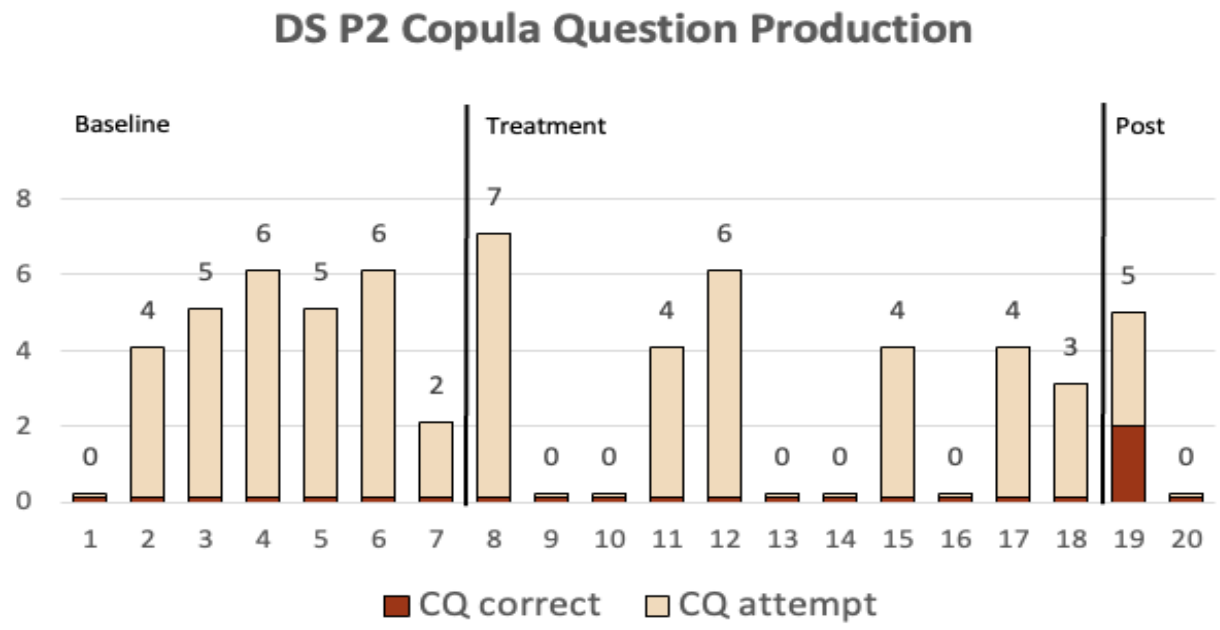

DS P1 Copula Declarative Production
Baseline
Treatment
Post
CD correct
CD attempt

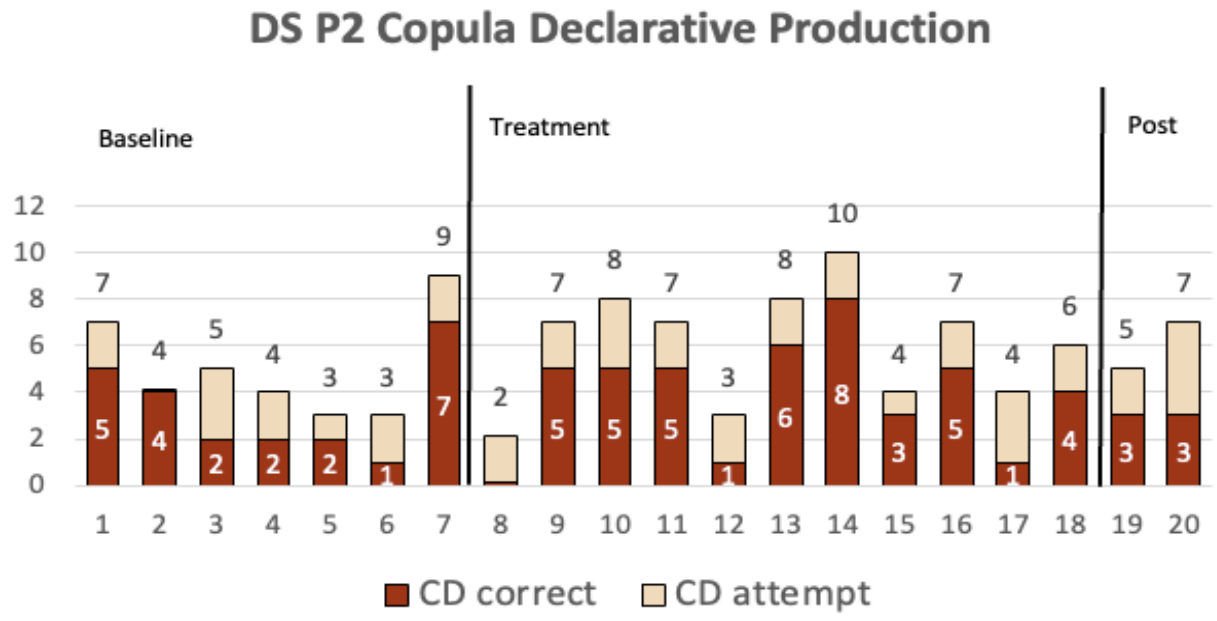
DS P2:


DS P2 Auxiliary Question Production
Baseline
Treatment
Post
0 0 1 10 11 9 4 9 2 0 10 8 3 0 9 3 9 7 7 7
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
AQ correct
AQ incorrect

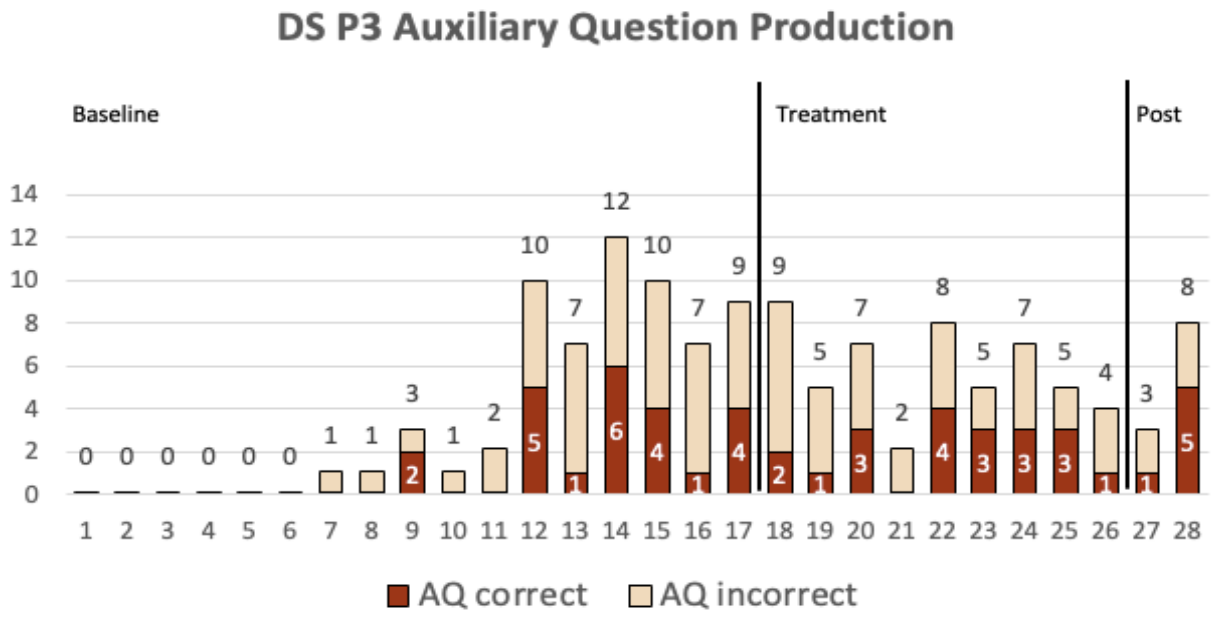

DS P2 Auxiliary Declarative Production
Baseline
Treatment
Post
16 14 15 7 8 9 11 9 13 20 6 10 17 18 11 17 11 10 10 14
7 4 8 2 5 7 5 5 9 15 5 6 10 13 6 12 10 5 5 8
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
AD correct
AD attempt

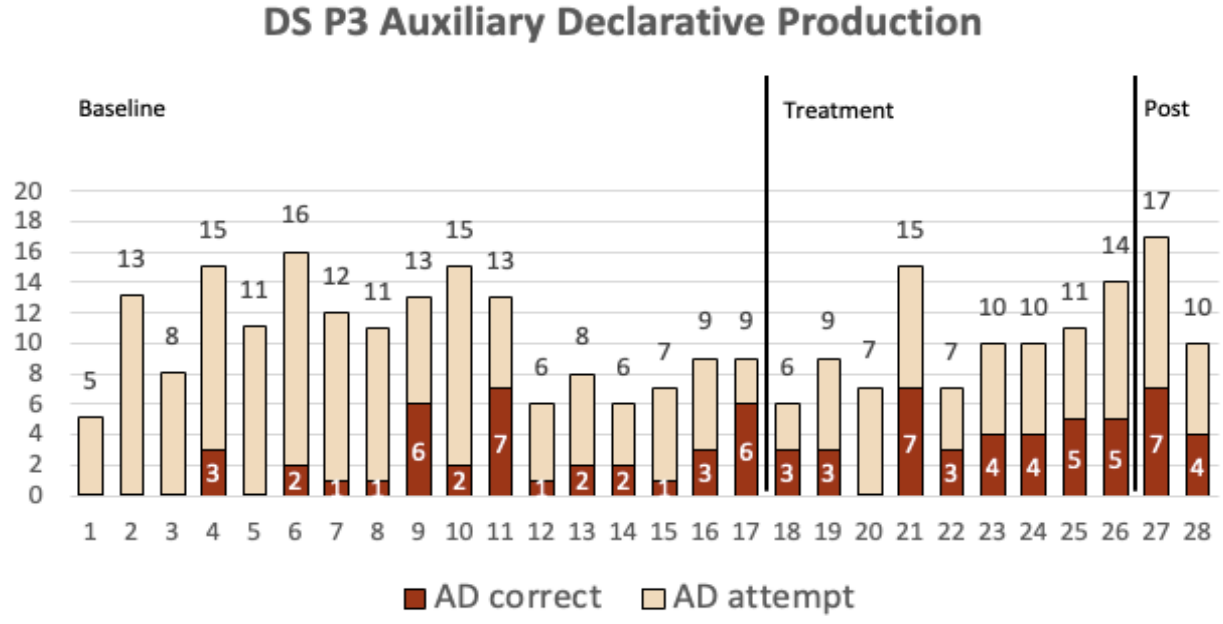
1012


DS P2 Copula Question Production
Baseline
Treatment
Post
0 4 5 6 5 6 2 7 0 0 4 6 0 0 4 0 4 3 5 0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
CQ correct
CQ attempt

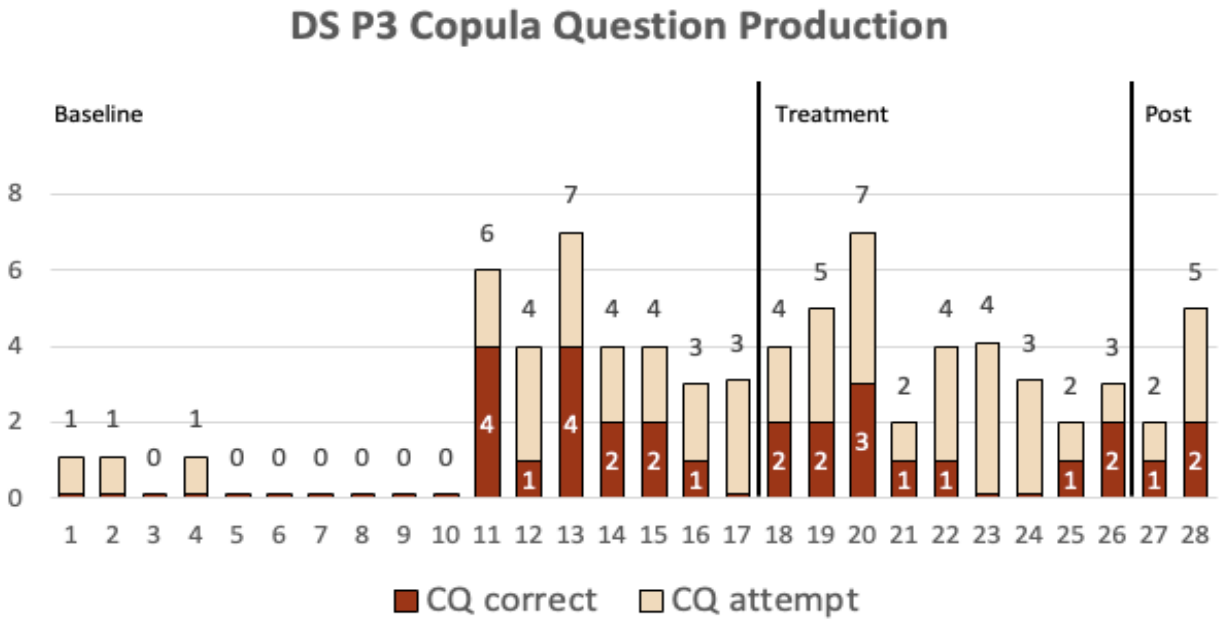

DS P2 Copula Declarative Production
Baseline
Treatment
Post
7 4 5 4 3 3 9 2 7 8 7 3 8 10 4 7 4 6 5 7
5 4 2 2 2 1 7 5 5 5 1 6 8 3 5 1 4 3 3
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
CD correct
CD attempt

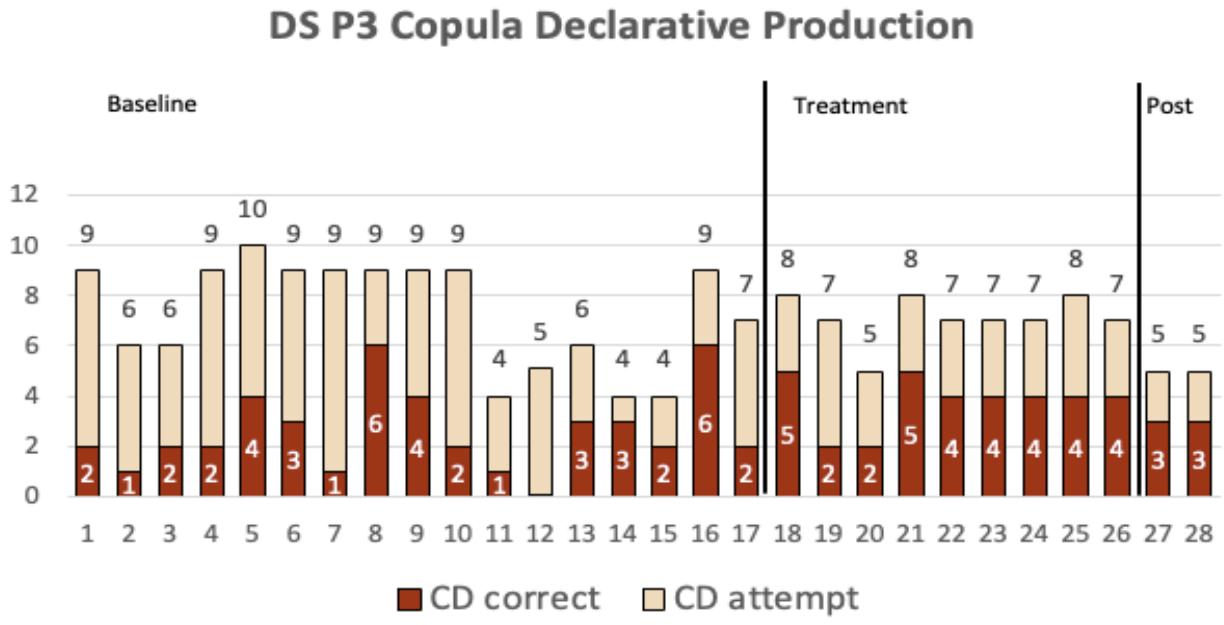
DS P3:


DS P3 Auxiliary Question Production
Baseline
Treatment
Post
AQ correct
AQ incorrect


DS P3 Auxiliary Declarative Production
Baseline
Treatment
Post
AD correct
AD attempt

1015
1016


DS P3 Copula Question Production
Baseline
Treatment
Post
CQ correct
CQ attempt


DS P3 Copula Declarative Production
Baseline
Treatment
Post
CD correct
CD attempt

Supplemental Material S4. Flowchart For Determining Continuation or Completion of Treatment.