Research Article

Targeting Complex Sentences in Older School Children With Specific Language Impairment: Results From an Early-Phase Treatment Study

Catherine H. Balthazar^a and Cheryl M. Scott^b

Purpose: This study investigated the effects of a complex sentence treatment at 2 dosage levels on language performance of 30 school-age children ages 10–14 years with specific language impairment.

Method: Three types of complex sentences (adverbial, object complement, relative) were taught in sequence in once or twice weekly dosage conditions. Outcome measures included sentence probes administered at baseline, treatment, and posttreatment phases and comparisons of pre–post performance on oral and written language tests and tasks. Relationships between pretest variables and treatment outcomes were also explored.

Results: Treatment was effective at improving performance on the sentence probes for the majority of participants; however, results differed by sentence type, with the largest

effect sizes for adverbial and relative clauses. Significant and clinically meaningful pre–post treatment gains were found on a comprehensive oral language test, but not on reading and writing measures. There was no treatment advantage for the higher dosage group. Several significant correlations indicated a relationship between lower pretest scores and higher outcome measures.

Conclusions: Results suggest that a focused intervention can produce improvements in complex sentence productions of older school children with language impairment. Future research should explore ways to maximize gains and extend impact to natural language contexts.

Supplemental Material: <https://doi.org/10.23641/asha.5923318>

A sizeable number of kindergarten children (7.4%) meet criteria for a specific language impairment (SLI), defined as weak language skills in spite of intact physical, social, and cognitive capabilities (Tomblin et al., 1997). For many, SLI is a persistent condition, extending into late elementary years and beyond (Conti-Ramsden, Botting, Simkin, & Knox, 2001; Tomblin, Zhang, Buckwalter, & O'Brien, 2003). A language impairment at school entry increases risk for reading and writing difficulties and takes a toll on academic achievement (Catts, Bridges, Little, & Tomblin, 2008).

A core feature of SLI is difficulty with syntax or grammar (Leonard, 2014). As preschoolers, English-speaking children with SLI struggle with the morphosyntactic features of language and inconsistently omit obligatory markers of verb tense, number, and aspect (Leonard & Deevy, 2010).

With age, additional grammatical issues that center on complex sentences emerge. A number of grammatical features contribute to sentence complexity; these include embedding/subordination, word order variations, and long-distance dependencies as found in *wh*-object questions or object relative clauses (Thompson & Shapiro, 2007). In comparison with age peers, there is considerable evidence that school-age children and adolescents with SLI find complex sentences more difficult to comprehend and produce (Fey, Catts, Proctor-Williams, Tomblin, & Zhang, 2004; Gillam & Johnston, 1992; Marinellie, 2004; Montgomery & Evans, 2009; Nippold, Mansfield, Billow, & Tomblin, 2008, 2009; Purdy, Leonard, Weber-Fox, & Kaganovich, 2014; Scott & Windsor, 2000). In this article, we are concerned with the first type of complexity: sentences with two clauses—one main clause and one subordinate clause.

To date, empirical reports of grammar-based treatments for older children with SLI are sparse, and even less common are studies that focus on complex sentences *per se*. Here, we report results from a treatment study designed to improve performance on complex sentences in school-age children with SLI ages 10 through 14 years. The protocol targets three types of multiclausal sentences: those containing

^aGovernors State University, University Park, IL

^bRush University, Chicago, IL

Correspondence to Catherine H. Balthazar: CBalthazar@govst.edu

Editor-in-Chief: Sean Redmond

Editor: Jan de Jong

Received March 21, 2017

Revision received August 1, 2017

Accepted November 28, 2017

https://doi.org/10.1044/2017_JSLHR-L-17-0105

Disclosure: The authors have declared that no competing interests existed at the time of publication.

a main clause and an adverbial, object complement, or relative subordinate clause. These three varieties account for the majority of multiclausal sentences that English-speaking children hear and produce (Diessel, 2004; Scott, 1988). We first present a review of studies showing that older children with SLI struggle with complex sentences—evidence that these types of sentences are worthy candidates for treatment. Next is a summary of empirical work focused on grammar objectives in this population. Finally, we present an overview of our study and conclude with three research questions.

Complex Sentences in SLI

Investigations of complex sentences in children with SLI fall into two broad categories. In one group of studies, researchers have probed listening comprehension in tasks requiring picture pointing or grammatical judgment. Researchers have also been interested in relationships between syntactic ability and reading comprehension. Other investigations have centered on the production of complex sentences in conversation, in narrative or expository spoken or written language samples, or in other elicited production tasks. Analyses have centered on the frequency, accuracy, and variety of complex sentence use in comparison to age and/or language peers.

Comprehension Studies

It should not be surprising that complex sentences pose comprehension problems for children with SLI. Sentence complexity often explains performance difficulties of adults and children with typical language ability in online and offline processing tasks (Chomsky, 1969; King & Just, 1991; Poirier & Shapiro, 2012). Beginning with Cromer (1978) and continuing to the present, researchers have reported that children with SLI have difficulty in comprehending sentences that violate canonical word order (e.g., reversible full passives) and those containing long distance dependencies (e.g., object *wh*-questions and object relatives) and other complex structures (cf. Friedmann & Novogrodsky, 2004; Montgomery & Evans, 2009; van der Lely, Jones, & Marshall, 2011). An investigation by Purdy and colleagues tested the influence of sentence complexity on listening comprehension in a direct way (Purdy et al., 2014). School-age children with and without SLI were asked to make grammatical judgments of verb agreement and finiteness errors in one-clause (simple) sentences (e.g., *every night they talks on the phone*) versus two-clause (complex) object complement sentences (*he makes the quiet boy talks a little louder*). Children with SLI were as accurate as age peers judging the correct/incorrect versions of verbs in simple sentences, but less accurate in complex sentences where syntactic constraints operated across two clauses. This study presents a convincing example that complex sentences “up the processing ante” for children with SLI. Of note, two of the three types of sentences targeted in our study, object complements and relatives, have figured prominently in listening comprehension studies.

Researchers have also pursued potential links between general syntactic ability and reading comprehension, finding

correlations between the two, as summarized by Scott and Koonce (2014). Although complex sentences as defined here have rarely been singled out as dependent variables in these investigations, it is probably not a coincidence that many students begin to struggle with reading comprehension in late elementary years when informational texts contain higher proportions of longer, complex sentences (Catts, Compton, Tomblin, & Bridges, 2012; Fang, 2012; Scott & Balthazar, 2010). Although our treatment protocol was not designed to impact reading comprehension per se, given the established links between syntactic ability and reading comprehension, we included a reading test among pre-post treatment measures.

Production Studies

Group comparisons of complex sentence use in children with and without language impairment have been reported in studies of spoken language samples (e.g., Marinellie, 2004; Nippold et al., 2008, 2009; To, Stokes, Cheung, & T'sou, 2010), in studies of written language samples (e.g., Dockrell, Lindsay, & Connelly, 2009; Suddarth, Plante, & Vance, 2012; Williams, Larkin, & Blaggan, 2013), and in studies that compare spoken and written samples (Gillam & Johnston, 1992; Scott & Windsor, 2000). Complex sentence usage typically has been indexed with global sentence-level metrics including mean sentence length (total words/total sentences), clause density (total clauses, main and subordinate/total sentences), and other measures such as percentage of complex sentences.

Group comparisons often find lower complex sentence frequencies in participants with language impairment (Marinellie, 2004; Morris & Crump, 1982; Nippold et al., 2008, 2009; Scott & Windsor, 2000; Strong, 1998). A study of narratives produced by Cantonese-speaking children found that a sentence complexity metric best distinguished SLI and typical language groups when compared with other sentence measures (To et al., 2010). A smaller group of studies have reported limited or no difference (Houck & Billingsley, 1989; Roth & Spekman, 1989; Zwislerlood, van Weerdenburg, Verhoeven, & Wijnen, 2015). It is important to note that genre, topic, and task are critical variables when comparing the effects of age and language ability on sentence complexity in naturalistic language samples (Nippold, 2010).

Several studies have provided more detailed analyses of specific types of complex sentences that distinguish SLI and typical language groups. Owen Van Horne and Lin (2011) reported that children with SLI produced fewer object complement complex sentences using lower-frequency cognitive verbs (e.g., *promise*, *decide*) in narrative/expository contexts. In Marinellie's (2004) work, school-age children with SLI used half the number of adverbial and relative clauses in conversation. In other work, relative clauses were the only type of complex sentence used at significantly lower rates in narratives told by 8-year-old children with SLI (Zwislerlood et al., 2015). Another distinguishing feature is the ability to combine several different types of subordinate clauses (e.g., a sentence containing both adverbial and

relative clauses; Gillam & Johnston, 1992; Marinellie, 2004). Scott and Lane (2008) reported that older children with SLI posted lower frequencies of (a) sentences with three or more clauses, (b) sentences with more than one level of subordination depth, and (c) sentences with relative clauses. Studies have also shown increases in error rates in complex sentences compared to simple sentences (e.g., Owen, 2010). Sentences with relative clauses are particularly error-prone (Frizelle & Fletcher, 2014; Jensen de Lopez, Sundahl Olsen, & Chrondrogianni, 2014; Novogrodsky & Friedman, 2006; Schuele & Nicholls, 2000). These more detailed analyses underscore that it is not only the absolute frequency of complex sentences but also the variety of subordinate clauses, the ways they are combined within a sentence, and the number of errors that distinguish sentences used by children with SLI.

Previous Treatment Studies Centered on Complex Sentences

A search for previous treatment studies that targeted complex sentences as we have defined them (sentences with a main and at least one subordinate clause) uncovered only two investigations with participants of comparable age to ours (ages 10–14). Hirschman (2000) targeted complex sentences with subordinate adverbial clauses in 9- and 10-year-old children enrolled in remedial classrooms (known to contain 80%–90% students with SLI). In 55 half hour, classroom-based sessions over the course of the school year, participants in the treatment group were (a) introduced to basic grammatical features of target sentences (e.g., subjects, objects, verbs, simple vs. complex sentences, subordinate conjunctions), (b) asked to identify target sentences in texts (Aesop's Fables), and (c) taught to rewrite fables presented as simple sentences using a sentence combining strategy. Hirschman reported significant gains for the treated group compared with the untreated group—a difference that held for 3 months after treatment concluded. Of note, participants with the lowest pretreatment complexity scores made more progress than those with higher scores.

In a second study directly applicable to our work, Levy and Friedmann (2009) worked with a 12-year-old boy on complex sentences with relative clauses, passive constructions, and *wh*-questions—all structures that involve long distance dependencies. Lessons (16 total) illustrated structural features (using color and shape codes and arrows) and provided guided practice constructing sentences. Pre–post treatment comparisons on a range of comprehension and elicited production tasks showed significant improvement that matched the performance of age peers in some cases.

A wider literature on sentence combining also relates to our work (as shown later, sentence combining was included among treatment activities). Sentence combining exercises teach students how to take two or more simple sentences and combine them into one longer sentence, which, depending on the nature of the input sentences, is often a complex, multiclausal sentence (Scott & Nelson, 2009). Participants in these studies have typically been general education students

and less frequently special education students. In a systematic review of 18 studies meeting inclusion criteria, Andrews et al. (2006) concluded that sentence combining provides an effective means of improving syntactic maturity in writing for students between the ages of 5 and 16. In many of these investigations, syntactic maturity has been indexed by measures of sentence length. In a meta-analysis of sentence combining and other writing instruction strategies, Graham and Perin (2007) reported a weighted mean effect size of .50 (moderate) for sentence combining across five investigations.

A smaller number of sentence combining studies have included participants with language impairments or weaknesses. Saddler and Graham (2005) randomly assigned fourth-grade writers to either sentence combining or traditional grammar instruction (e.g., teaching parts of speech) groups. Although there were no group differences in quality ratings of first draft writing, the sentence combining group performed better on posttreatment tasks and also on their revision ability. Similar to the work of Hirschman (2000) cited above, the treatment had its greatest impact on students with the lowest pretreatment language scores. In a later study, Cantonese-speaking school children with language impairment were randomly assigned to sentence combining or narrative-based syntactic treatment where both treatments targeted the same complex sentences (To, Lui, Li, & Lam, 2015). Positive and comparable effects were seen for both protocols. In a review of studies using a sentence combining protocol for children with language impairment, Datchic and Kabina (2013) described gains in percentages of complex sentences and general writing quality using experimenter-created rubrics.

We have restricted this review of treatment studies to those specifically targeting complex sentences. A summary of the effectiveness of a broader group of grammar-based intervention studies for school-age children with primary language impairments uncovered 35 studies categorized according to specific grammatical targets, age of participants, method, delivery, results, degree of experimental control, and other variables (Ebbels, 2014). Of these, 25 studies targeted specific syntactic structures, most commonly morphosyntax targets including verb and noun inflections, verb auxiliaries (9), questions (5), passives (5), argument structure (4), and pronouns (3). Only one study (Levy & Friedmann, 2009, reviewed above) targeted complex sentences. Ten of the 35 studies targeted improvements in expressive language but were not target-specific. Only eight (of 35) studies included children within the age range of participants in our study (ages 10–14). By including three major types of complex sentences in a treatment protocol delivered to older school-age children, our study addresses a neglected structural domain of grammatical targets in a neglected age group of students with language impairment.

Study Purpose and Design

This study examines the effects of a language treatment designed to build understanding and use of complex sentences in children with SLI between the ages of 10 and

14. The age range of 10–14 was chosen for several reasons. This period corresponds with continuing developmental growth in the use of complex sentences (Nippold, Hesketh, Duthie, & Mansfield, 2005; Nippold, Mansfield, & Billow, 2007; To et al., 2010) as well as academic language skills generally (Uccelli et al., 2015), thus avoiding potential ceiling effects. Furthermore, this is the time when a sizeable number of children are identified as late-emergent poor readers whose reading comprehension difficulties are often related to general language comprehension issues, including sentence comprehension (Catts et al., 2012). We have discussed in previous reports the specific grammatical features of complex sentences found in academic texts that would challenge individuals with language impairment in this age range (Scott & Balthazar, 2010, 2013). We also wished to address the substantial gap in the evidence base for language treatments with older school-age children and adolescents. A range of 4 years was considered adequate to reveal any obvious effects of age on treatment results.

Treatment targets included each of the three types of complex sentences (adverbial, object complement, relative) that, together, account for a majority of English complex sentences investigated in developmental studies and in studies of children with language disorders (Scott, 1988). Targeting all three varieties of sentences is, to our knowledge, unique to this study where we hoped to determine if treatment effects varied by sentence type. Although adverbial, object complement, and relative clauses share the status of being subordinate to a main clause (i.e., dependent), they are structurally quite different (see Appendix), and we suspected that relative clauses may be more difficult to treat, given their late development, low frequency, and error susceptibility (Scott & Lane, 2008; Zwitserlood et al., 2015). The three sentence types were treated consecutively for equal amounts of time with identical procedures that included exposure and repetition, identification, and scaffolded manipulation activities in one-on-one sessions. The protocol was designed to be consistent with studies showing that treatment effects are increased by using a greater variety of target stimuli (Plante et al., 2014) and by using deductive (in our terminology, metalinguistic) activities (Finestack & Fey, 2009). In order to examine effects of treatment intensity, we compared two intensity levels commonly found in the language treatment literature. Specific research questions were as follows:

1. Are there significant treatment effects on three different types of two-clause sentences (those with one subordinate adverbial, object complement, or relative clause) on a sentence-level probe task, and does increased treatment intensity produce greater effects? We hypothesized that treatment would be effective for all three sentence types and that the higher dosage group would demonstrate greater gains.
2. Are there significant pre–post gains to oral language test scores, reading test scores, and writing sample measures, and does increased treatment intensity produce greater gains? We examined changes in scores on sentence-level language tasks and reading comprehension tasks, as well as indicators of sentence complexity in written tasks, to explore the potential scope of impact of the treatment.
3. Does the age, nonverbal cognitive score, or any pre-treatment language score of a participant correspond to the degree of treatment-induced change? We explored participant factors that might mitigate response to treatment, as well as pretest scores that could help us identify those most likely to benefit from treatment of complex sentences. We suspected that age and cognitive ability would be important, given the continued growth in complex sentence knowledge and the expression of more complex concepts and ideas throughout adolescence.

Method

Participants

Participants were selected from a pool of children between the ages of 10;0 and 14;11 (years;months) who had been referred to the study by speech-language pathologists and parents because they demonstrated problems with reading comprehension, writing composition, oral discourse, and listening or following directions and had profiles consistent with a diagnosis of SLI. We collected case history documents including a parent questionnaire and whenever possible the current Individualized Education Plan for each child referred. We administered tests to confirm nonverbal intelligence scores within the average range and oral language test scores below the average range. We also administered the first of three pretest complex sentence probes. This initial information was evaluated to confirm that participants met inclusionary and exclusionary criteria for SLI and that their test performance included evidence of errors formulating and/or understanding complex sentences. Of the 47 potential participants evaluated, 30 met criteria for study participation: (a) a current score of at least 80 on the Test of Nonverbal Intelligence–Third and Fourth Editions (TONI-3 [Brown, Sherbenou, & Johnsen, 1997] and TONI-4 [Brown, Sherbenou, & Johnsen, 2010]), (b) hearing within normal limits as determined by pure-tone audiometric screening administered within 6 months of study enrollment, (c) language performance 1 *SD* or more below the mean on the Core Language Quotient (Core) of the Clinical Evaluation of Language Fundamentals–Fourth Edition (CELF-4; Semel, Wiig, & Secord, 2003), (d) no other diagnosed conditions or disorders known to affect language performance, (e) less than 66% accuracy on at least one sentence type on the experimental probe task designed for this study, and (f) English as a primary language. For five participants, parents reported a secondary language spoken in the home (Spanish, Chinese, Hindi, or Cantonese). Three participants identified as Hispanic, seven as African American or Black, three as Asian, and four as “Eurasian” or “African Caucasian.” The remaining participants identified as non-Hispanic, White, monolingual speakers of English.

Participants were assigned to a treatment dosage schedule of once (1/wk) or twice (2/wk) weekly sessions. In most cases, assignment to a treatment dosage group was based on feasibility (i.e., scheduling constraints for the participant and clinician). When both once and twice weekly sessions were feasible, the first author assigned participants to a group in order to achieve balance between the two groups with respect to age and gender.

Fourteen participants were assigned to the 1/wk level, ranging in age from 10 years to 13 years 10 months, with a mean of 11 years 6 months. Sixteen participants were assigned to the 2/wk level, ranging in age from 10 years to 14 years 11 months, with a mean of 12 years 1 month. Two-tailed independent samples *t* tests were completed in order to evaluate potential differences between the two groups in age, nonverbal intelligence, or oral language scores. There were no significant differences between the dosage groups for age ($p = .398$), nonverbal intelligence ($p = .753$), or CELF-4 Core Language Score ($p = .717$). Table 1 provides a summary of the gender, age, nonverbal intelligence, and oral language scores of the participants.

Clinicians

A total of 10 clinicians provided the pre-post testing and treatment at four university-affiliated clinics and four public schools. All clinicians were certified and licensed speech-language pathologists or graduate students in speech-language pathology supervised on site by certified and licensed speech-language pathologists. Study clinicians included the two authors, two graduate student assistants working under the supervision of the authors, one doctoral fellow, and five practicing clinicians. All clinicians completed a training protocol, described in the section below on treatment fidelity.

Outcome Measures

Outcome measures were of two types: (a) complex sentence probes administered before, during, and after treatment and (b) language tests and criterion referenced tasks administered pre-post treatment and chosen to sample a broad array of language skills including oral and written language. These measures are described below.

Table 1. Participant summary.

Group	<i>n</i>	M	F		Age (years;months)	TONI	CELF-4
1/wk	14	8	6	Mean	11;6	94.27	73.73
				Range	10;0–13;10	81–127	48–87
2/wk	16	10	6	Mean	12;1	92.67	71.19
				Range	10;0–14;11	81–118	54–85
Total	30	18	12		11;9	92.71	72.10

Note. M = male; F = female; TONI = standard score on Test of Nonverbal Intelligence–Third or Fourth Edition; CELF-4 = Core Language Quotient of the Clinical Evaluation of Language Fundamentals–Fourth Edition.

Complex Sentence Probes

The experimental probe was a written sentence combining task developed to sample production of target complex sentences in a format that resembled one of the treatment activities. Each probe item contained two single-clause sentences that were to be combined into a complex sentence, given a starter word or phrase. The items were balanced to represent content similar to sentences found in either narrative/conversational or informational/expository discourse, similar to sentences used in treatment sessions. Probe items also matched structural variations in the three types of complex sentences taught (see Appendix). Adverbial clause (AC) probes contained six items (three left-branching and three right-branching forms); object complement (OC) probes contained six items (three finite and three nonfinite forms); relative clause (RC) probes contained eight items (two each of the four possible combinations of subject or object modification with either subject or object relativization).¹ Participants wrote their responses; however, typing or dictation were permitted variations.

Pre-Post Measures

As a broad measure of oral language ability, the CELF-4 Core (Semel et al., 2003) subtests were administered pre- and posttreatment. Three of the four subtests, Formulated Sentences (FS), Recalling Sentences (RS), and Concepts and Following Directions (CFD), were sentence-level tasks whose items included complex sentences with adverbial, relative, and object complement subordinate clauses. The Sentence Comprehension (SC) subtest of the Comprehensive Assessment of Spoken Language (CASL; Carrow-Woolfolk, 1999) provided an additional measure of complex sentence comprehension. The Gray Oral Reading Tests–Fourth Edition (GORT-4; Wiederholt & Bryant, 2001) Oral Reading Quotient was included as a broad measure of reading fluency and comprehension.

Written language production was sampled in two ways. Narrative writing was elicited using the Story Construction subtest of the Test of Written Language–Fourth Edition (TOWL-4; Hammill & Larsen, 2009). Expository writing was elicited by asking participants to summarize a short video (NOVA ScienceNow) following the procedure developed by Scott and Windsor (2000). Two different videos on the same topic (the nature of memory) were used for pre- and posttesting. Both narrative and expository samples were analyzed for two measures commonly used to index sentence complexity: mean length of T-unit (MLTU) in words and the subordination index (SI; number of clauses, main and subordinate, divided by number of T-units).

Scoring

Experimental probe items were scored by the second author using a 0–3 point scoring rubric developed for this

¹Examples of experimental probe items are available in Supplemental Material S1.

study.² The highest score was given to a grammatical, complex sentence that was appropriate given the content and form of the stimulus. The lowest score was given to incomplete or unchanged sentences. Scores in between were assigned when a response met some but not all criteria. Percent accuracy scores were then calculated for each sentence type (AC, OC, RC).³ Scoring of the CELF-4, the CASL SC subtest, and the GORT-4 was completed by the clinician who administered the test. Written samples were transcribed by the second author into language analysis software (Systematic Analysis of Language Transcripts, Miller & Iglesias, 2012) using T-unit segmentation rules (Hunt, 1970) and coded for the presence of subordinate clauses. MLTU and SI measures were calculated automatically by the Systematic Analysis of Language Transcripts software.

Reliability

After initial scoring, the first author rescored a subset of probes and written samples for reliability. Participant and treatment phase (pre or post) identifiers were first removed from all probes and samples. Fifteen percent of the probes were then randomly selected and rescored. Interrater reliability was calculated through point-by-point analysis. The total number of agreements on each probe was divided by the number of sentences scored to arrive at a percent agreement score. These were averaged over the 580 sentences for a total of 90% exact agreement between the two authors. This level of interrater agreement was deemed adequate to support the use of the scoring rubric for reliably quantifying important features of participant responses. Ten percent of written language samples were selected for reliability checks of T-unit segmentation and SI coding. Segmentation agreement was 99%. Point-by-point coding agreement for subordinate clauses was 91.4%. Where disagreements existed, the probe scores and written sample codes assigned by the second author were used.

Spoken responses on the CELF-4 and GORT-4 were audio-recorded and reviewed by trained graduate assistants who did not participate in the testing or treatment. The recordings were compared against the handwritten notes on test protocols for verification. All of the test protocols were then rescored by the same graduate assistants, and any discrepancies were brought to the attention of the first author, who scored the protocol again to determine a final score. In all, there were five test protocols subjected to rescoring by the first author. Together, these reliability procedures and analyses were designed to provide a measure of control against experimenter bias.

²The scoring rubric is available in Supplemental Material S2.

³We use the acronyms AC, RC, and OC to refer to complex sentences with two clauses—one main clause and one subordinate clause, where the subordinate clause is either an adverbial (AC), relative (RC), or object complement clause (OC).

Complex Sentence Treatment

Treatment Procedures

Treatment sessions were 40–60 min in length and followed a fixed sequence of activities provided in the treatment manual. The activities were presented in both oral and written form using computer-based applications (PowerPoint and Word) and paper-and-pencil tasks. Written forms of stimuli were always available as a reference in order to reduce load on working memory. Based on participant preference and computer skills, clinicians and participants decided together which version (computer or paper/pencil) to use for each activity. All presentations of the target sentence types included visual emphasis of key sentence features with boldface type, color, highlighting, and underlining. All complex sentences were composed of two clauses: a main clause and one subordinate clause. All items were read aloud by the clinician. Participants were instructed to visually track the sentence as it was read aloud and thus had simultaneous visual and auditory input.

All complex sentences used in treatment were drawn from a pool of sentences developed specifically for this project by the authors and trained graduate assistants. Sentences provided exemplars of the structural variations addressed under each complex sentence type (e.g., both right- and left-branching AC; see variations of AC, OC, and RC in the Appendix). The sentences were written to reflect narrative or expository content. Narrative content centered on everyday life occurrences (e.g., *John could not leave for the movies until his mother got home from work*). Expository content was based on information adapted from upper elementary social studies and science textbooks (e.g., *After the bombing of Pearl Harbor, President Franklin D. Roosevelt declared war on Japan*). Sentences used in each treatment session were then balanced across structural and content domains. Sentences in each treatment session were unique exemplars; none were repeated in subsequent sessions or in the complex sentence probes.

Each session began with a “warm-up”—a brief presentation of the structural characteristics of the targeted sentence type for that session. Instruction was scripted and included a definition of the target clause, how it could be used, and ways to identify and differentiate it from other clause types. The clinician then read aloud with the participant a one- or two-paragraph selection containing multiple examples of the target sentence type to show how the complex sentence worked in context. Next, the clinician modeled five complex sentences for repetition (*Sentence Repetition*) for the purpose of priming production of the desired structural patterns.

The second part of each session was devoted to explicit metalinguistic instruction and guided practice. The first of these activities was *Sentence Identification*. This included five sentences for which the participant identified the target subordinate clause after three to five examples provided by the clinician. The second metalinguistic activity alternated between *Sentence Deconstruction* and *Sentence Combining* or *Sentence Generation*. Following three items that were

reviewed jointly with the clinician, five items were presented for the participant to attempt independently. Scaffolded support (a cue, prompt, or explanation) was offered when the participant was unable to respond correctly independently. In *Sentence Deconstruction*, participants were shown a complex sentence and asked to isolate the subordinate clause from the main clause. In *Sentence Combining*, participants were asked to combine two simple sentences into one complex sentence. For structural reasons, *Sentence Combining* was altered slightly into a *Sentence Generation* task in OC sessions. For both of these tasks, the goal was for the participant to use predetermined simple (one-clause) sentences to construct complex (two-clause) sentences that followed the target syntactic pattern.

The third part of each session engaged the participant in one of several activities designed to highlight the meanings and/or purposes of the complex sentence in context. For the *Clause Hunt*, the participant and clinician read a passage aloud together. After each two or three sentences, the clinician prompted the participant to identify target subordinate clauses, followed by discussion on how the clause added meaning, detail, or nuance to the sentence. For *Preference Production*, participants viewed sets of two complex sentences that were the same with the exception of one key word or phrase in the subordinate clause (e.g., *Some soldiers who are leaving for duty in Iraq are having trouble finding a job* versus *Some soldiers who are returning from duty in Iraq are having trouble finding a job*). The participant was asked which one made more sense and why. In the *Cloze Production* activity, the participant and clinician read a passage aloud together. On a second copy of the same passage with blanks for subordinate clauses, the participant attempted to provide a target clause that made sense. In all of these contextualized activities, the clinician provided scaffolding as needed.

Probes were administered at the end of the treatment activities within a designated session. For each probe set, the clinician began by reading a script that illustrated how to complete the sentence combining task. Once these instructions were completed, the clinician read each item (two simple sentences) aloud while the participant read along silently. The participant then wrote his or her solution (combining the information from two simple sentences into one complex sentence) in the blank provided.⁴

Treatment Fidelity

Clinicians were trained to deliver the treatment and were provided with a scripted protocol and materials. Training included information about AC, OC, and RC structure and scaffolding techniques. A treatment manual provided schedules and scripts for all activities. Clinicians reviewed the manual with the trainer, viewed videotaped examples of treatment sessions, and engaged in role-play demonstrations of each treatment activity. Project staff remained in close

contact by phone and onsite visits to respond to clinician questions.

In this manner, a highly scripted protocol was used to guide every treatment session and probe administration. Each clinician worked from the same master schedule that laid out the sequence of pretesting (two to three sessions), treatment (nine or 18 sessions, depending on dosage assignment), and post testing (two to three sessions). The manualized nature of the treatment in terms of both scripting and scheduling served to maximize treatment fidelity.

Clinicians recorded the number of items completed in each session. The resulting data allowed for comparison of dose frequency and adherence to the specified schedule of stimuli and activities across participants. In 82% of sessions, the clinician documented the specified number of trials of sentence-level activities, including sentence identification, sentence deconstruction, and sentence combining/generation.

Dosage

The concept of “dosage” encompasses multiple factors that affect how much treatment is delivered (Warren, Fey, & Yoder, 2007). In this study, total dosage was varied by manipulating dose frequency (how often the treatment was delivered), whereas dose, dose form, and total intervention duration were held constant. Participants were assigned to one of two dosage levels. For both levels, the average rate of stimulus presentation per session remained constant at 30 per session. Session length (40–60 min) and total duration of intervention (9 weeks) were also the same. Stimulus presentation included 15 instances of modeling and repetition and a target of 15 trials in which participants were asked to manipulate a complex sentence with clinician feedback/scaffolding. The 1/wk group attended once-weekly sessions for 9 weeks, for a total of nine sessions (three sessions each for AC, OC, and RC, presented consecutively). The 2/wk group attended twice-weekly sessions for 9 weeks, for a total of 18 sessions (six sessions each for AC, OC, and RC). The cumulative intervention intensity for the 2/wk condition was therefore double that of the 1/wk session. Once treatment was completed, we calculated the actual dosage delivered to each group. The difference in cumulative intervention intensity was very close to what was planned. Cumulative intervention intensity for the 1/wk condition was 236 total stimulus items delivered on average (26 items per session for nine sessions). This compared with 502 total stimulus items for the 2/wk condition (28 items per session for 18 sessions).

Design

Single-Subject Experimental Design

A key element of the study was to determine whether the treatment was equally effective across the three sentence types. We employed a single-subject experimental design for this purpose. A multiple baseline across behaviors design allowed us to provide the treatment to all participants non-concurrently over a 3-year time period. The multiple baseline design establishes within-subject control over extraneous

⁴Sample experimental probe items are available in Supplemental Material S1.

variables such as history and practice effects by comparing several developmentally related target behaviors—in our case, three types of complex sentences—during baseline and sequentially arranged treatment phases. A causal link between the treatment and changes in the outcome measure (the complex sentence probe) can then be inferred when a pattern of stability in baseline followed by improvement during treatment is repeated across multiple targets (Kratochwill et al., 2010; Robey, Schultz, Crawford, & Sinner, 1999). The design consisted of five phases, each comprising a predetermined number of sessions (see Figure 1).

Pretreatment (Phase A1). Participants referred to the study were first evaluated to establish eligibility. In this initial session, we collected parent consent/participant assent, a brief history from the parent and/or school clinician, and completed oral language (CELF-4 Core), nonverbal intelligence (TONI-3/TONI-4) testing, and a complex sentence probe. Those who qualified for the study then participated in two additional pretreatment evaluation sessions, during which sentence comprehension (CASL SC), reading (GORT-4), narrative writing (TOWL-4 Story Construction), and expository writing (Written Summary) were assessed. Probes were administered at each of the three pretreatment sessions to establish baseline performance.

Treatment (Phases B1, B2, B3). Participants in the 1/wk group met individually with a clinician once per week for 9 weeks. They received three sessions each targeting AC, OC, and RC, in that order. During the treatment period for each sentence type, probes were administered

in the session prior to treatment, at the end of the second treatment session, in the first session posttreatment, and 1 week posttreatment. Participants in the 2/wk group met individually with a clinician twice per week for 9 weeks. They received six sessions each targeting AC, OC, and RC, in that order. During the treatment period, target sentence probes were administered for each target sentence type immediately prior to treatment, at the end of the second, fourth, and sixth treatment sessions, and 1 week posttreatment. RC sentences, which were the last to be treated, were also probed once during the third OC treatment session.

Posttreatment (Phase A2). Three sessions of evaluation began 1 week after the final treatment session and were completed within 4 weeks of the end of the treatment sequence. All pretreatment language and reading tests and writing tasks were once again administered. Final probes of all three sentence types were administered during each of the posttreatment evaluation sessions. By the end of the final posttreatment session, a total of eight probe sets for the 1/wk group and 11 probe sets for the 2/wk group were collected for each complex sentence type.

Data Analysis

Single-Subject Data Analysis

Probe data from each of the 30 participants were analyzed to establish the magnitude of a treatment effect for each sentence type. We combined visual analysis with two quantitative measures, percent exceeding median (PEM)

Figure 1. Phases of the multiple baseline across behaviors design, showing test administration schedule. For each sentence type, treatment phase is indicated in gray. There were three treatment sessions with one probe during treatment for the 1/wk participants and six treatment sessions with two probes during treatment for the 2/wk participants. A1 = pretest phase; B1 = treatment phase, first target: AC; B2 = treatment phase, second target: OC; B3 = treatment phase, third target: RC; A2 = posttest phase; TONI = Test of Nonverbal Intelligence—Third or Fourth Edition (Brown, Sherbenou & Johnsen, 1997, 2010); CELF-4 = Clinical Evaluation of Language Fundamentals—Fourth Edition (Semel, Wiig, & Secord, 2003); CASL = Sentence Comprehension subtest of the Comprehensive Assessment of Spoken Language (Carrow-Woolfolk, 1999); GORT-4 = Gray Oral Language Test—Fourth Edition Oral Reading Quotient (Wiederholdt & Bryant, 2001); WS = Written Summarization; TOWL-4 Story = Story Generation subtest of the Test of Written Language—Fourth Edition (Hammill & Larsen, 2009); AC = adverbial clauses; OC = object complements; RC = relative clauses.

Pre-Test			Treatment			Post-Test		
TONI4	GORT4	TOWL4	AC Treated, Probe (B1)	AC Probe (A2)			CELF4	GORT4
CELF4	WS	CASL	OC Probe (A1)	OC Treated, Probe (B2)	OC Probe (A2)		TOWL4	WS
AC, OC, RC Probe (A1)	AC, OC, RC Probe (A1)	AC, OC, RC Probe (A1)	RC Probe (A1)	RC Probe (A1)	RC Treated, Probe (B3)	RC Probe (A2)	AC, OC, RC Probe (A2)	CASL

and standard mean difference, pooled (SMD_p). Visual analysis provided observation of the slope of trend lines during baseline in order to identify participants whose performance may have been improving prior to treatment, as could be the case due to factors such as practice effects or regression to the mean. A baseline was considered to be stable if the last pretreatment probe measure remained at or below the median of baseline measures. PEM helped us quantify performance improvements during treatment by observing how often a participant's performance exceeded the baseline median. The number of data points above the baseline median during and after treatment was divided by the total number of postbaseline data points to arrive at the PEM. Although PEM does not capture the magnitude or variability of performance, it has the advantage of reflecting treatment effects even in the presence of floor or ceiling data. Following Wendt (2009), we reasoned that PEM scores of .90 or more would be expected of highly effective treatment. Moderately effective treatment would be indicated by PEM scores less than .90 but more than .70, and PEM scores of less than .70 would indicate questionable or ineffective treatment.

SMD_p was calculated to provide a standardized measure of the amount of change attributable to treatment. Given the relatively small number of baseline data points, we reasoned that SMD_p , a method that uses all data points to arrive at the means and variances, would be the best model for representing expected variance (Ebert & Kohnert, 2009). We calculated SMD_p by subtracting the mean of the baseline data points from the mean of all data points following the introduction of treatment and dividing by the pooled standard deviation across all data points (Busk & Serlin, 1992). In the absence of effect size data from comparable treatment studies, we ranked SMD_p effect sizes for each sentence type across all 30 participants and used quartile values to define minimum values for small, medium, and large effect sizes (25th, 50th, and 75th percentiles, respectively). To analyze dosage effects, a two-way analysis of variance was completed with dosage (1/wk, 2/wk) and sentence type (AC, OC, RC) as independent variables and SMD_p or PEM as dependent variables.

Analysis of Pre–Post Data

Pre–post comparisons between the two dosage groups on oral and written language outcome measures (Research Question 2) were completed using a repeated-measures multivariate analysis of variance (MANOVA) with dosage (1/wk, 2/wk) as the between-subjects factor and time (Pre, Post) as the within-subject factor. Complex sentence probe averages pretreatment (Phase A1) and posttreatment (Phase A2) were also used in this analysis (see Figure 1). In order to evaluate effect size, partial eta squared (η^2) values were calculated. In a MANOVA, partial eta squared expresses the proportion of variance in the dependent variables accounted for by the independent variables (dosage or time).

Analysis of Participant Factors

To explore possible factors related to treatment response (Research Question 3), data from the two dosage

groups were combined. Pretreatment chronological age, non-verbal intelligence (TONI-3/TONI-4), oral language ability (CELF-4 Core and subtests), reading ability (GORT-4), and complex sentence performance (pretest probe scores for AC, OC, and RC) were entered into a correlational analysis with two indicators of treatment outcomes: (a) SMD_p as a measure of effect size on probe performance and (b) amount of pre–post gain on all other measures. Because numerous comparisons were being explored in the correlation, we adjusted the alpha level to .03 using the false discovery rate procedure (Benjamini & Hochberg, 1995) in order to set a stringent criterion for significance.

Results

Single-Subject Results

Treatment Effects

Visual analysis indicated that baselines for the large majority of participants were stable prior to treatment for all three types of sentences (93% for AC, 69% for OC, and 86% for RC). A PEM score was calculated for each participant on each sentence type, which was then used to determine whether treatment had been highly (better than .90) or moderately (better than .70) effective for each sentence type. Means and standard deviations for PEM are presented in Table 2. On average, moderately effective PEM values were obtained for AC (.73) and RC (.80), but not OC (.49). Twenty-six participants (87%) demonstrated moderately to highly effective PEM scores on at least one of the sentence types. Ten (33%) of the participants achieved a moderate to highly effective PEM on two sentence types; six participants (20%) achieved it on all three. Four participants (13%) did not demonstrate moderately or highly effective PEM scores for any of the sentence types.

Standardized effect size values (SMD_p) were calculated for each participant on each sentence type (see Table 2).⁵ Effect size values are reported in Table 3. For the OC sentences, a number of participants demonstrated no or negative treatment change, and consequently, the first quartile value fell below zero. Because it would be problematic to interpret negative values as improvement, we set the lower boundary value for a small OC effect size at zero rather than the 25th percentile. Medium (.15) and large (.55) effect sizes for OC were still defined using the 50th and 75th percentile values. Twenty-four participants (80%) demonstrated a medium or large effect size on at least one of the sentence types. Seven (23%) of the participants achieved a medium or large effect size on two sentence types; eight participants (27%) achieved it on all three.

Dosage and Sentence Type Effects

In order to detect differences in treatment effect size among the two dosage groups and three sentence types, data were subjected to a two-way analysis of variance with

⁵Benchmark values for magnitude of SMD_p may be found in Supplemental Material S3 for this article.

Table 2. Mean (standard deviation) PEM and SMD_p values for each sentence type.

Measure	Group	AC	OC	RC
PEM	1/wk	0.81 (0.31)	0.37 (0.36)	0.81 (0.31)
	2/wk	0.68 (0.28)	0.60 (.23)	0.79 (0.23)
	Total	0.73 (0.30)	0.49 (.32)	0.80 (0.27)
SMD _p	1/wk	1.12 (0.74)	-0.09 (0.85)	0.89 (0.76)
	2/wk	0.71 (0.28)	0.20 (0.45)	1.10 (0.54)
	Total	0.91 (0.69)	0.06 (0.67)	1.00 (0.65)

Note. PEM = percent exceeding median; SMD_p = standard mean difference using pooled variance; AC = adverbial clauses; OC = object complement clauses; RC = relative clauses.

sentence type (AC, OC, RC) and dosage (1/wk, 2/wk) as the independent variables and PEM or SMD_p as the dependent variable. Means and standard deviations for each dosage group by sentence type are presented in Table 3. Mean values reflected a significant effect of Sentence Type for both PEM, $F(2, 27) = 22.863, p < .001$, and SMD_p, $F(2, 27) = 32.458, p < .001$. The source of difference was significantly lower performance on OC compared to the other two sentence types. Surprisingly, although the number of treatment sessions was doubled for the 2/wk participants, we found no significant main effect of dosage and no significant Dosage $\times$ Sentence Type interaction on SMD_p for AC, $F(1) = 2.801, p = .105$; OC, $F(1) = 1.391, p = .248$; or RC, $F(1) = 0.743, p = .396$. There was also no dosage or Dosage $\times$ Sentence Type interaction for PEM for AC, $F(1) = 1.234, p = .277$, and RC, $F(1) = 0.067, p = .797$; however, there was a statistically significant Dosage $\times$ Sentence Type interaction found for OC, $F(1) = 4.401, p = .045$.

Analysis of Pre-Post Test Results

Pre and post probe scores, CELF-4 scores, CASL, and GORT-4 scores are presented in Table 4. For the language and reading test scores, results of the Time $\times$ Dosage repeated-measures MANOVA revealed a significant effect of time overall, $F(9, 20) = 9.825, p < .001, \Lambda = .184, \eta^2 = .816$. Univariate tests showed that the effect of time was significant for the CELF-4 Core, $F(1) = 41.05, p < .001, \eta^2 = .59$; CFD, $F(1) = 9.53, p = .005, \eta^2 = .30$; FS, $F(1) = 46.59, p < .001, \eta^2 = .63$; and word classes (WC), $F(1) = 10.36, p = .003, \eta^2 = .27$, measures, as well as the pre-post sentence probes for AC, $F(1) = 32.56, p < .001, \eta^2 = .54$; OC, $F(1) = 4.25, p = .049, \eta^2 = .13$; and RC, $F(1) = 46.75, p < .001, \eta^2 = .63$. The effect of time approached significance for SC ($p = .058$) and GORT-4 ($p = .071$). It was not significant for RS ($p = .347$) or word definitions (WD; $p = .379$). The effect of dosage overall was not significant, $F(9, 20) = 1.22, p = .339, \Lambda = .184$. The Time $\times$ Dosage interaction was also not significant, $F(9, 20) = 1.80, p = .131, \Lambda = .552$.

Writing sample scores (MLTU and SI) are presented in Table 5. For the writing scores, results of the Time $\times$ Dosage repeated-measures MANOVA revealed no significant effect of time overall for the TOWL-4 Story Generation,

$F(6, 23) = .487, p = .811, \Lambda = .887$, and no Time $\times$ Dosage interaction, $F(6, 23) = .873, p = .53, \Lambda = .815$. For the expository writing samples (WS), the effect of time overall was significant, $F(6, 20) = 4.862, p = .003, \Lambda = .407$, as was the Time $\times$ Dosage interaction, $F(6, 20) = 2.657, p = .046, \Lambda = .556$, but not in the anticipated direction. Univariate tests showed that these effects were due to a significant effect of time on SI for both groups overall, $F(1) = 6.658, p = .016$, produced by a decrease in SI from pretest ($M = 2.44$) to posttest ($M = 2.03$) and of the Time $\times$ Dosage interaction on MLTU, $F(1) = 4.973, p = .035$, which resulted from a drop in MLTU on the WS from pretest ($M = 12.21$) to posttest ($M = 9.48$) for the 2/wk group.

Analysis of Participant Factors

The pretest variables of chronological age, TONI-3/TONI-4 score, CELF-4 Core score, CELF-4 subtest scores, CASL SC score, GORT-4 score, and average baseline AC, OC, and RC probe scores were entered into a correlational analysis with effect sizes (SMD_p values) and pre-post gain values on the CELF-4 Core and subtests, GORT-4, and CASL. Five of 49 possible correlations were significant, all involving sentence-level measures in which lower pretest scores correlated with greater treatment effects and gains (i.e., negative correlations).⁶ The FS pretest score was correlated with OC effect size ($r^2 = -.400, p = .029$) and FS gain ($r^2 = -.394, p = .030$). AC average at baseline produced a significant correlation with AC effect size ($r^2 = -.628, p = .000$). OC average at baseline correlated with OC effect size ($r^2 = -.569, p = .001$). Age, CELF-4 Core, TONI-3/TONI-4, CASL, and GORT-4 pretest scores did not correlate significantly with effect sizes or test gains.

Discussion

This study was designed as an early-phase treatment study addressing whether school-age children with SLI benefit from a treatment protocol designed to improve performance on three types of complex sentences. Benefit was examined by analyzing probe scores as part of a single-subject multiple baseline across behaviors design and comparing pre-post treatment measures that included norm-referenced language tests and naturalistic writing samples. We also examined relationships between outcome measures and participant age, nonverbal cognition, and pretreatment language scores.

Treatment Effects

Our primary research question was whether there was a significant effect of treatment on participants' production of each of the three types of complex sentences, and if so, whether the effect was greater in the higher dosage condition. Performance on complex sentence probes varied across

⁶A table showing all correlations is available in Supplemental Material S4.

Table 3. Mean (standard deviation) and gain scores for AC, OC, and RC probes.

Group	AC			OC			RC		
	Pre	Post	Gain	Pre	Post	Gain	Pre	Post	Gain
1/wk	.63 (.20)	.84 (.08)	.21 (.19)	.73 (.13)	.76 (.11)	.03 (.17)	.38 (.22)	.58 (.17)	.20 (.17)
2/wk	.71 (.10)	.81 (.10)	.10 (.11)	.67 (.15)	.75 (.10)	.08 (.13)	.34 (.18)	.59 (.23)	.25 (.19)
Total	.67 (.15)	.82 (.09)	.15 (.16)	.70 (.14)	.76 (.13)	.06 (.15)	.36 (.20)	.58 (.20)	.22 (.18)

Note. Pretest score is average of first three baseline probes. Posttest score is average of last three posttreatment probes. AC = adverbial clauses; OC = object complement clauses; RC = relative clauses.

individuals in terms of the size of treatment effect (expressed as the PEM and SMD_p values) and which sentence type(s) improved but did not vary as a function of treatment dosage.

The first measure of treatment effect was PEM, which reflected the proportion of measurements above the median baseline after treatment. Performance for most participants improved once treatment began and remained reliably above baseline levels, thus demonstrating clinical significance on at least one and usually more than one sentence type. The second way we looked at treatment effect was with SMD_p , a standardized effect size metric that quantifies the magnitude of change. The majority of participants (80%) demonstrated medium or large effect sizes for at least one of the three sentence types; a sizable proportion also demonstrated medium to large effects for two (23%) or all three (27%) sentence types. We concluded that the majority of participants made important treatment-related gains in complex sentence production.

Effect size analyses revealed a difference in treatment effect across the three types of sentences. The majority of participants demonstrated moderate or high learning effects on AC and RC sentences, but not on OC sentences. Compared with AC and RC, OC performance showed greater fluctuation across time and was also closer to ceiling at baseline, which could have limited the potential for contrast.

Differences in the syntactic operations underlying performance on OC probes could also explain treatment effect disparities. AC and RC probes involved conjoining, substitution, and movement operations different from those required for OC probes. The OC is an obligatory argument (object) of the main clause, which would be ungrammatical without it. Also, only certain main clause verbs, ones that signal cognitive or linguistic activity (e.g., *think*, *ask*, *decide*, *conclude*), collocate with OC. These structural and semantic features of OC sentences required that participants follow a different metalinguistic pathway during probes and in treatment activities.

Pre-Post Comparisons

To explore the broader potential impact of treatment beyond direct probe measures, we compared pre-post oral language and reading test scores as well as sentence-level indicators of complexity in writing samples. Across dosage groups, the average gain on the CELF-4 Core was large—a statistically significant 10 standard score points. There are several reasons to interpret this change as attributable to treatment and clinically significant and not the result of regression toward the mean or practice effects. First is the size of the improvement—the large number of participants (60%) whose gain exceeded the 90% confidence interval of

Table 4. CELF-4 Core and subtests, CASL SC, and GORT-4 mean (standard deviation) and gain scores by dosage group.

Test	1/wk			2/wk			Total		
	Pre	Post	Gain	Pre	Post	Gain	Pre	Post	Gain
CELF-4	73.1 (9.8)	82.9 (9.8)	9.8 (6.8)	71.2 (11.7)	80.7 (14.4)	9.5 (9.2)	72.1 (10.7)	81.7 (12.3)	9.6 (8.1)
CFD ^a	5.3 (2.2)	6.3 (3.1)	1.0 (2.0)	3.2 (1.1)	5.3 (3.2)	1.9 (2.5)	4.2 (2.0)	5.8 (3.1)	1.6 (2.2)
WD ^{b,c}	9.0 (0.0)	10.0 (1.4)	1.0 (1.4)	6.8 (3.2)	8.6 (3.2)	.75 (2.2)	7.5 (2.7)	9.0 (2.8)	.83 (1.8)
RS	3.9 (1.8)	4.2 (2.1)	0.3 (1.4)	4.0 (2.4)	4.2 (2.9)	0.2 (1.7)	3.9 (2.1)	4.2 (2.5)	0.3 (1.5)
FS	6.1 (2.5)	9.4 (2.4)	3.3 (2.4)	6.4 (3.6)	9.3 (3.7)	2.9 (2.5)	6.2 (3.1)	9.3 (3.1)	3.1 (2.5)
WC	6.5 (1.8)	8.1 (2.3)	1.6 (2.2)	6.6 (1.7)	7.6 (2.5)	1.0 (2.2)	6.5 (1.7)	7.8 (2.3)	1.3 (2.2)
CASL SC	84 (8)	87 (13)	3 (12)	83 (7)	87 (8)	4 (7)	83 (7)	87 (11)	4 (10)
GORT-4	81 (12)	84 (15)	3 (10)	72 (15)	77 (13)	5 (11)	77 (14)	80 (14)	4 (10)

Note. CELF-4 = Core Language Quotient of the Clinical Evaluation of Language Fundamentals–Fourth Edition; CFD = Concepts and Following Directions; WD = Word Definitions; RS = Recalling Sentences; FS = Formulated Sentences; WC = Word Classes; CASL SC = Sentence Comprehension subtest of the Comprehensive Assessment of Spoken Language; GORT-4 = Gray Oral Language Test–Fourth Edition Oral Reading Quotient.

^a n = 12 in the 1/wk group, 12 in the 2/wk group. ^b n = 2 in the 1/wk group, 4 in the 2/wk group at pretest. ^c n = 2 in the 1/wk group, 5 in the 2/wk group at posttest.

Table 5. Pre- and posttest narrative and expository writing sample score mean (standard deviation) scores by dosage group.

Measure	Group	TOWL-4 Story		Written Summary	
		Pre	Post	Pre	Post
MLTU	1/wk	9.84 (2.98)	8.87 (2.77)	11.62 (4.88)	11.86 (5.19)
	2/wk	9.00 (2.73)	9.13 (2.21)	12.21 (3.07)	9.48 (1.89)
	Total	9.39 (2.83)	9.02 (2.45)	11.93 (3.85)	10.36 (3.89)
SI	1/wk	2.82 (0.470)	1.65 (0.498)	2.45 (0.953)	2.16 (0.845)
	2/wk	1.67 (0.440)	1.77 (0.453)	2.43 (0.799)	1.92 (0.523)
	Total	1.74 (0.450)	1.72 (0.471)	2.39 (0.839)	1.99 (0.696)

Note. TOWL-4 Story = Story Generation subtest of the Test of Written Language—Fourth Edition (Hammill & Larsen, 2009); MLTU = mean length of T-unit (the number of total words divided by the total number of T-units); SI = subordination index (the number of total clauses divided by the total number of T-units).

the test (± 6 points), an interval calculated to set statistical boundaries on the likelihood that a given test score is stable and representative. In research on the longitudinal course of language impairment, it is reported that school-age children rarely exceed 1 *SE* of measurement unit on repeated test administrations (this point is reviewed in Gillam et al., 2008). Second is the number of participants (47%) whose score normalized from a clinical status of SLI (a standard score of ≤ 85) to a score above 85. A third indication is the variation in improvement across CELF-4 Core subtests (see Table 4). If regression toward the mean, practice effects, or the passage of time underlies the improvement, one might expect gains to be evenly distributed across subtests. However, there were no significant gains on two subtests—RS and Word Definitions. Furthermore, of the four subtests that make up the CELF-4 Core, it was FS, the subtest most similar to treatment activities, which posted the highest gains and contributed the most to the Core Language Quotient gains.

We found less evidence of treatment impact on reading and writing scores. Although both dosage groups demonstrated reading gains approaching significance on the GORT-4, the size of the change did not reach statistical significance or exceed the test's confidence interval (± 6). Because passages on the GORT-4 contain complex sentences that must be correctly parsed to arrive at text-level meaning, we had hypothesized that scores could be affected by our treatment. Because the changes observed were consistently in a positive direction, it may be that either the quantity of treatment was not sufficient to produce significant gains or that the treatment content did not link directly enough to the task of reading.

Unlike reading, we saw no improvements on the most naturalistic of our pre-post measures—the two writing tasks. Posttreatment narrative and expository writing samples did not show gains on measures frequently used to index sentence complexity (MLTU, SI). The lack of significant positive change in either narrative or expository writing was not surprising given the short overall length of treatment (9 weeks, with a total of nine or 18 treatment sessions), the sentence-level emphasis of the intervention,

and the relative difficulty of writing for school-age children with SLI (Scott & Windsor, 2000).

Dosage

We examined whether the 2/wk dosage produced significant differences in treatment effects or pre-post gains. We did not find an advantage to this doubling of the treatment intensity on any of our oral and written language outcome measures. A negative finding regarding a dosage effect is somewhat counterintuitive, and it is important to consider why it may have occurred. It could be that performance approached ceiling for both dosage groups after the first few sessions (there were three sessions devoted to each sentence type in the 1/wk group), and there was no advantage to additional treatment. This seems likely when looking at findings from a study using the same probes with typically developing 10-, 12-, and 14-year-olds (Selin, 2013). In this study, a plateau in performance emerged at around the age of 12, with scores ranging from 60% (RC) to 90% (AC and OC) accuracy for 14-year-olds. In terms of average age and postprobe scores, our participants were similar (average age of 12, RC score of 58%, and AC score of 82%). Thus, many of the participants, even within the lower dosage group, were approaching levels of performance comparable to that of age peers. A probe task with greater discriminatory ability at the higher end may be needed in order to observe treatment dosage effects.

Pre-post gains on the CELF-4 were significant, but again, not greater for the higher dosage group. A ceiling effect would not account for the lack of a dosage effect on this measure. We suspect that the impact on the CELF-4 Core score resulted from treatment effects on the FS subtest, which, as we stated previously, is composed of a high proportion of test items involving AC constructions. Perhaps a norm-referenced measure, which relied in equal proportions on OC and RC constructions, would have more potential to show additional impact. In our review of potential measures, however, no such test was found. For the remaining pre-post measures, namely the reading and writing tasks, significant gains were not observed for either of the dosage groups. Here it is likely that even the higher dosage was

not enough to significantly affect performance on more complex tasks, such as reading and writing.

Other language treatment studies across a variety of language domains and involving both adults and children have similarly found little relationship between dosage and degree of treatment response (e.g., Barratt, Littlejohns, & Thomson, 1991; Fey, Warren, Yoder, & Bredin-Oja, 2013; McGregor, Sheng, & Ball, 2007). Our findings, along with results in these studies, highlight the challenges of defining the concept of dosage in the study of language treatment (Proctor-Williams, 2009) and, more specifically, understanding the role of treatment intensity (Yoder, Fey, & Warren, 2012). Further investigation is required to confirm that there is no added benefit for the higher dosage and to explore potential dosage effects and interactions with longer total treatment times, treatment maintenance, and different outcome measures.

Pretest Characteristics and Outcomes

Our last research question involved an exploration of factors that could potentially mediate treatment effects, including baseline probe measures, age, nonverbal cognition, and pretest language scores. Results were surprising in that so few of these variables correlated with treatment outcomes. Only a small subset of baseline probe and pretest language scores corresponded with the amount of treatment gain, and for these, the correlation was a negative one. Lower AC and OC probes at pretest corresponded to larger effect sizes, suggesting that participants who had the most difficulties with these complex sentences made the greatest gains. This pattern is consistent with results from the two studies that taught complex sentence structure to students in the same age range as our participants. Hirschman (2000) found that students with lower complexity scores showed greater benefit from the treatment of adverbial clauses. Similarly, Saddler and Graham (2005) noted that students with the lowest pretreatment language scores benefitted the most from sentence combining instruction. We took this pattern, combined with the significant effect size for RC (for which pretest performance was lower than AC and OC to begin with), as an encouraging sign that the treatment protocol was working to improve exactly the skills intended, for those who needed it the most. It also suggests that a fairly narrow focus on complex sentences during assessment is required in order to determine which individuals might be in need of support, specifically on complex sentence structure. Likewise, the only language measure that related significantly (and negatively) to effect size was FS. As discussed above, this is also a sentence-level production task with many exemplars of AC sentences, one of the complex sentence types treated.

Future Directions

Our analysis of the results leaves us optimistic about providing complex sentence intervention to older school-age children with SLI. The study has provided a starting point

to build a stronger evidence base supporting treatment decisions and valuable lessons that will influence future investigations. Regarding methodology, future studies should extend assessments posttreatment to determine the degree to which learning is maintained. In evaluating the dosage component of the study, randomization of assignment to dosage levels would improve our ability to make general, more conclusive statements about dosage effects. Later-phase iterations of complex sentence treatment studies can consider further methodological improvements that enhance internal and external validity.

Specific to treatment, there are a number of avenues to explore. The nonsignificant dosage effect raises the question of whether our dosage increment was sufficient. Although it should not be assumed that “more equals better” (Kamhi, 2014), it may be that a larger difference between the two dosage levels in either frequency (number of sessions per week) or cumulative intervention intensity (total amount of instruction provided) would have made an observable difference. Going back to the early days of language treatment research, clinicians have discussed how much treatment is needed for change to occur, as well as the appropriate performance criteria for language goals (Fey, 1986). Future investigations should address these issues. Regarding content, several modifications based on our results deserve consideration. Treatment could pay more attention to meaning, for example, emphasizing complement-taking verbs, particularly lower-frequency verbs found in school subjects (e.g., *hypothesize*, *assume*), or logical relations signaled by various adverbial conjunctions. The number and variety of clauses being manipulated in each sentence and the depth of embedding, both of which increase the difficulty level for parsing a complex sentence, could also be increased. It would be interesting to see whether more treatment time devoted to manipulating target sentences in real reading and writing tasks might promote generalization. These remain areas worthy of further investigation, and we look forward to determining whether changes might result in broader effects that are maintained in more natural language contexts. As an early-phase treatment study, however, this study has demonstrated that older school-age children can be taught to be more successful manipulating untrained complex sentences. Furthermore, there are clinically meaningful gains on a comprehensive test of oral language, and these changes can occur in a relatively short intervention span. The results support a developing narrative wherein older children and adolescents with SLI benefit from highly focused interventions that recruit specific language skills and address their persistent functional language problems.

Acknowledgments

The Complex Sentence Intervention treatment protocol was developed with the support of Grant 1R15011165-01 from the National Institute on Deafness and Other Communicative Disorders, awarded to Catherine H. Balthazar and Cheryl M. Scott. The content of this article is solely the responsibility of the authors

and does not necessarily represent the official views of the National Institute on Deafness and Other Communicative Disorders or the National Institutes of Health. The authors wish to acknowledge the speech-language pathologists and graduate students who contributed to the development and delivery of the treatment protocol: Laura Anderson, Joy Bedell, Patty Boyd, Julie Burns, Sara Butler, Erica Fenton, Cynthia Loiterman, Robyn Maciejewski, Colleen Shanahan, Katie Stuepfert, Pat Tattersall, and, especially, Nicole Koonce and Claire Selin.

References

- Andrews, R., Torgerson, C., Beverton, S., Freeman, A., Locke, T., Law, G., ... Zhu, D. (2006). The effect of grammar teaching on writing development. *British Educational Research Journal*, 32, 39–55.
- Barratt, J., Littlejohns, P., & Thompson, J. (1991). Trial of intensive compared with weekly speech therapy in preschool children. *Archives of Disease in Childhood*, 67, 106–108.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, 57(1), 289–300.
- Brown, L., Sherbenou, R. J., & Johnsen, S. K. (1997). *Test of Non-verbal Intelligence—Third Edition*. Austin, TX: Pro-Ed.
- Brown, L., Sherbenou, R. J., & Johnsen, S. K. (2010). *Test of Non-verbal Intelligence—Fourth Edition*. Austin, TX: Pro-Ed.
- Busk, P. L., & Serlin, R. C. (1992). Meta-analysis for single-case research. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case research design and analysis: New directions for psychology and education* (pp. 187–212). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Carrow-Woolfolk, E. (1999). *Comprehensive Assessment of Spoken Language*. Circle Pines, MN: AGS Publishing.
- Catts, H. W., Bridges, M. S., Little, T. D., & Tomblin, J. B. (2008). Reading achievement growth in children with language impairments. *Journal of Speech, Language, and Hearing Research*, 51, 1569–1579.
- Catts, H. W., Compton, D., Tomblin, J. B., & Bridges, M. S. (2012). Prevalence and nature of late-emerging poor readers. *Journal of Educational Psychology*, 104(1), 166–181.
- Chomsky, C. (1969). *The acquisition of syntax in children from 5 to 10*. Cambridge, MA: MIT Press.
- Conti-Ramsden, G., Botting, N., Simkin, Z., & Knox, E. (2001). Follow-up of children attending infant language units: Outcomes at 11 years of age. *International Journal of Language & Communication Disorders*, 36, 207–219.
- Cromer, R. F. (1978). The basis of childhood dysphasia: A linguistic approach. In M. A. Wyke (Ed.), *Developmental dysphasia* (pp. 85–134). New York, NY: Academic Press.
- Datchic, S. M., & Kabina, R. M. (2013). A review of teaching sentence-level writing skills to students with writing difficulties and learning disabilities. *Remedial and Special Education*, 34(3), 180–192.
- Diessel, H. (2004). *The acquisition of complex sentences*. Cambridge, United Kingdom: Cambridge University Press.
- Dockrell, J. E., Lindsay, G., & Connelly, V. (2009). The impact of specific language impairment on adolescent written text. *Exceptional Children*, 75(4), 427–446.
- Ebbels, S. H. (2014). Effectiveness of intervention for grammar in school-aged children with primary language impairments: A review of the evidence. *Child Language Teaching and Therapy*, 30(1), 7–40.
- Ebert, K. D., & Kohnert, K. (2009). Nonlinguistic cognitive treatment for primary language impairment. *Clinical Linguistics & Phonetics*, 23, 647–654.
- Fang, Z. (2012). Language correlates of disciplinary literacy. *Topics in Language Disorders*, 32(1), 19–34.
- Fey, M. (1986). *Language intervention with young children*. Needham Heights, MA: Allyn & Bacon.
- Fey, M., Catts, H., Proctor-Williams, K., Tomblin, J. B., & Zhang, X. (2004). Oral and written story composition skills of children with language impairment. *Journal of Speech, Language, and Hearing Research*, 47, 1301–1318.
- Fey, M., Warren, S., Yoder, P., & Bredin-Oja, S. (2013). Is more better? Milieu communication teaching in toddlers with intellectual disabilities. *Journal of Speech, Language, and Hearing Research*, 56, 679–693. NIHMS493548.
- Finestack, L. H., & Fey, M. E. (2009). Evaluation of a deductive procedure to teach grammatical inflections to children with language impairment. *American Journal of Speech-Language Pathology*, 18, 289–302.
- Friedmann, N., & Novogrodsky, R. (2004). The acquisition of relative clause comprehension in Hebrew: A study of SLI and normal development. *Journal of Child Language*, 31, 661–681.
- Frizelle, P., & Fletcher, P. (2014). Relative clause constructions in children with specific language impairment. *International Journal of Communication Disorders*, 49, 255–264.
- Gillam, R., & Johnston, J. (1992). Spoken and written language relationships in language/learning impaired and normally achieving school-age children. *Journal of Speech and Hearing Research*, 35, 1303–1315.
- Gillam, R., Loeb, D. F., Hoffman, L. M., Bohman, R., Chaplain, C. A., Thibodeau, L., & Friel-Patti, S. (2008). The efficacy of Fast ForWord language intervention in school-age children with language impairment: A randomized controlled trial. *Journal of Speech, Language, and Hearing Research*, 51, 97–119.
- Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, 99, 445–476.
- Hammill, D., & Larsen, S. (2009). *Test of Written Language—Fourth Edition*. Austin, TX: Pro-Ed.
- Hirschman, M. (2000). Language repair via metalinguistic means. *International Journal of Language & Communication Disorders*, 35, 251–268.
- Houck, C., & Billingsley, B. (1989). Written expression of students with and without learning disabilities. *Journal of Learning Disabilities*, 22, 561–568.
- Hunt, K. W. (1970). Syntactic maturity in school children and adults. *Monographs of the Society for Research in Child Development*, 35(1), iii–iv, 1–67.
- Jensen de Lopez, K., Sundahl Olsen, L., & Chronrogianni, V. (2014). Annoying Danish relatives: Comprehension and production of relative clauses by Danish children with and without SLI. *Journal of Child Language*, 41, 51–83.
- Kamhi, A. (2014). Improving clinical practices for children with language and learning disorders. *Language, Speech, and Hearing Services in Schools*, 45, 92–103.
- King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, 30(5), 580–602.
- 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*. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf
- Leonard, L. B. (2014). *Children with specific language impairment* (2nd ed.). Cambridge, MA: MIT Press.
- Leonard, L. B., & Deevy, P. (2010). Tense and aspect interpretation in children with specific language impairment. *Journal of Child Language*, 37, 385–418.

- Levy, H., & Friedmann, N. (2009). Treatment of syntactic movement in syntactic SLI: A case study. *First Language*, 29, 15–50.
- Marinellie, S. (2004). Complex syntax used by schoolage children with specific language impairment (SLI) in child-adult conversation. *Journal of Communication Disorders*, 37, 517–533.
- McGregor, K. K., Sheng, L., & Ball, T. (2007). Complexities of expressive word learning over time. *Language, Speech, and Hearing Services in Schools*, 38(4), 353–364.
- Miller, J., & Iglesias, A. (2012). Systematic Analysis of Language Transcripts (Research Version 2012) [Computer software]. Middleton, WI: SALT Software, LLC.
- Montgomery, J. W., & Evans, J. L. (2009). Complex sentence comprehension and working memory in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 52, 269–288.
- Morris, N., & Crump, D. (1982). Syntactic and vocabulary development in the written language of learning disabled and non-learning disabled students at four age levels. *Learning Disability Quarterly*, 5, 163–172.
- Nippold, M. A. (2010). Explaining complex matters: How knowledge of a domain drives language. In M. A. Nippold & C. M. Scott (Eds.), *Expository discourse in children, adolescents, and adults* (pp. 41–61). New York, NY: Psychology Press.
- Nippold, M. A., Hesketh, L. J., Duthie, J. K., & Mansfield, T. C. (2005). Conversational versus expository discourse: A study of syntactic development in children, adolescents, and adults. *Journal of Speech, Language, and Hearing Research*, 48, 1048–1064.
- Nippold, M. A., Mansfield, T. C., & Billow, J. L. (2007). Peer conflict explanations in children, adolescents, and adults: Examining the development of complex syntax. *American Journal of Speech-Language Pathology*, 16, 179–188.
- Nippold, M. A., Mansfield, T. C., Billow, J. L., & Tomblin, J. B. (2008). Expository discourse in adolescents with language impairments: Examining syntactic development. *American Journal of Speech-Language Pathology*, 17, 356–366.
- Nippold, M. A., Mansfield, T. C., Billow, J. L., & Tomblin, J. B. (2009). Syntactic development in adolescents with language impairments: A follow-up investigation. *American Journal of Speech-Language Pathology*, 18, 241–251.
- Novogrodsky, R., & Friedman, N. (2006). The production of relative clauses in syntactic SLI: A window to the nature of impairment. *International Journal of Speech-Language Pathology*, 8(4), 364–375.
- Owen, A. J. (2010). Factors affecting accuracy of past tense production in children with specific language impairment and their typically developing peers: The influence of verb transitivity, clause location, and sentence type. *Journal of Speech, Language, and Hearing Research*, 53, 993–1014.
- Owen van Horne, A. J., & Lin, S. (2011). Cognitive state verbs and complement clauses in children with SLI and their typically developing peers. *Clinical Linguistics and Phonetics*, 25(10), 881–898.
- Plante, E., Ogilvie, T., Vance, R., Aquilar, J., Dailey, N., Meyers, C., . . . Burton, R. (2014). Variability in the language input to children enhances learning in a treatment context. *American Journal of Speech-Language Pathology*, 23, 530–545.
- Poirier, J., & Shapiro, L. (2012). Linguistic and psycholinguistic foundations. In R. Peach & L. Shapiro (Eds.), *Cognition and acquired language disorders: An information processing approach* (pp. 121–146). St. Louis, MO: Elsevier Mosby.
- Proctor-Williams, K. (2009). Dosage and distribution in morphosyntax intervention. *Topics in Language Disorders*, 29, 294–311.
- Purdy, J. D., Leonard, L. B., Weber-Fox, C., & Kaganovich, N. (2014). Decreased sensitivity to long-distance dependencies in children with a history of specific language impairment: Electrophysiological evidence. *Journal of Speech, Language, and Hearing Research*, 57, 1049–1059.
- Robey, R., Schultz, M., Crawford, A., & Sinner, C. (1999). Single subject clinical-outcome research: Designs, data, effect sizes, and analyses. *Aphasiology*, 13(6), 445–473.
- Roth, F., & Spekman, N. (1989). The oral syntactic proficiency of learning disabled students: A spontaneous story sampling analysis. *Journal of Speech and Hearing Research*, 32, 67–77.
- Saddler, B., & Graham, S. (2005). The effects of a peer-assisted sentence-combining instruction on the writing performance of more and less skilled young writers. *Journal of Educational Psychology*, 97, 43–54.
- Schuele, M. C., & Nicholls, L. M. (2000). Relative clauses: Evidence of continued linguistic vulnerability in children with specific language impairment. *Clinical Linguistics and Phonetics*, 14(8), 563–585.
- Scott, C. (1988). Spoken and written syntax. In M. Nippold (Ed.), *Later language development: Ages nine through nineteen* (pp. 49–95). San Diego, CA: College-Hill Press.
- Scott, C., & Balthazar, C. (2010). The grammar of information: Challenges for older students with language impairments. *Topics in Language Disorders*, 30(4), 288–307.
- Scott, C., & Koonce, N. (2014). Syntactic contributions to literacy learning (pp. 283–301). In A. Stone, E. Silliman, B. Ehren, & G. Wallach (Eds.), *Handbook of language and literacy: Development and disorders* (2nd ed.). New York, NY: Guilford.
- Scott, C., & Lane, S. (2008, June). Capturing sentence complexity of school-age children with/without language impairment. Poster presented at the Symposium for Research in Child Language Disorders, Madison, WI.
- Scott, C., & Nelson, N. (2009). Sentence combining: Assessment and intervention applications. *Language and Learning Education*, 16, 14–20.
- Scott, C., & Windsor, J. (2000). General language performance measures in spoken and written narrative and expository discourse in schoolage children with language learning disabilities. *Journal of Speech, Language, and Hearing Research*, 43, 324–339.
- Scott, C. M., & Balthazar, C. (2013). The role of complex sentence knowledge in children with reading and writing difficulties. *Perspectives on Language and Literacy*, 39(2), 16–24.
- Selin, C. (2013). *The effect of age on a forced sentence completion task in children ten to fourteen*. Unpublished master's thesis, Rush University, Chicago, IL.
- Semel, E., Wiig, E., & Secord, W. (2003). *Clinical Evaluation of Language Fundamentals—Fourth Edition*. San Antonio, TX: Harcourt.
- Strong, C. (1998). *Strong narrative assessment procedure*. Eau Claire, WI: Thinking Publications.
- Suddarth, R., Plante, E., & Vance, R. (2012). Written narrative characteristics in adults with language impairment. *Journal of Speech, Language, and Hearing Research*, 55, 409–420.
- Thompson, C., & Shapiro, L. (2007). Complexity in treatment of syntactic deficits. *American Journal of Speech-Language Pathology*, 18, 30–42.
- To, C. K. S., Lui, H. M., Li, X. X., & Lam, G. Y. H. (2015). A randomized controlled trial of two syntactic treatments with Cantonese-speaking school-age children with language disorders. *Journal of Speech, Language, and Hearing Research*, 58, 1258–1272.

- To, C. K. S., Stokes, S., Cheung, H.-T., & T'sou, B. (2010). Narrative assessment for Cantonese-speaking children. *Journal of Speech, Language, and Hearing Research*, 53, 648–669.
- Tomblin, J. B., Records, N., 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, 1245–1260.
- Tomblin, J. B., Zhang, X., Buckwalter, P., & O'Brien, M. (2003). The stability of primary language disorder: Four years after kindergarten diagnosis. *Journal of Speech, Language, and Hearing Research*, 46, 1283–1296.
- Uccelli, P., Barr, C. D., Dobbs, C. L., Galloway, E. P., Meneses, A., & Sanchez, E. (2015). Core academic language skills (CALS): An expanded operational construct and a novel instrument to chart school-relevant language proficiency in preadolescent and adolescent learners. *Applied Psycholinguistics*, 36(5), 1077–1109. <https://doi.org/10.1017/S014271641400006X>
- van der Lely, H., Jones, M., & Marshall, C. R. (2011). Who did Buzz see someone? Grammaticality judgements of wh-questions in typically developing children and children with Grammatical-SLI. *Lingua*, 121, 408–422.
- Warren, S. F., Fey, M. E., & Yoder, P. J. (2007). Differential treatment intensity research: A missing link to creating optimally effective communication interventions. *Mental Retardation and Developmental Disabilities Research Reviews*, 13, 70–77.
- Wendt, O. (2009). Research on the use of manual signs and graphic symbols in autism spectrum disorders: A systematic review. In P. Mirenda & T. Iacono (Eds.), *Autism spectrum disorders and AAC* (pp. 83–140). Baltimore, MD: Paul D. Brookes.
- Wiederholt, J. L., & Bryant, B. R. (2001). *Gray Oral Reading Tests—Fourth Edition*. Austin, TX: Pro-Ed.
- Williams, G. J., Larkin, R. F., & Blaggan, S. (2013). Written language skills in children with specific language impairment. *International Journal of Communication Disorders*, 48, 160–171.
- Yoder, P. J., Fey, M. E., & Warren, S. F. (2012). Studying the impact of intensity is important but complicated. *International Journal of Speech Language Pathology*, 14(5), 410–413.
- Zwitserslood, R., van Weerdenburg, M., Verhoeven, L., & Wijnen, F. (2015). Development of morphosyntactic accuracy and grammatical complexity in Dutch school-age children with SLI. *Journal of Speech, Language, and Hearing Research*, 58, 891–905.

Appendix

Examples of Sentence Targets (AC, OC, RC) and Structural Variations

Complex sentences with adverbial clauses (AC): Adverbial clauses modify the associated main clause, adding information about time, place, or manner/condition. They are adjoined to the main clause with an adverbial conjunction (such as *because*, *when*, *although*, *unless*). The AC can either follow (right-branching) or precede (left-branching) the main clause.

- *Right-branching:* Matthew couldn't get his books because his locker was jammed shut. (N)⁷
- *Left-branching:* Although the Titanic was unsinkable, it went down in the North Atlantic on April 15, 1912. (E)

Complex sentences with object complement clauses (OC): Object complement clauses are embedded in the main clause as the obligatory object of the verb. A restricted set of main clause verbs take object complements; these include cognitive verbs (e.g., *think*, *know*, *say*, *conclude*, *decide*, *predict*). OC take a variety of forms including finite forms beginning with *that* or a wh-interrogative form (*what*, *why*) and nonfinite forms in which the verb in the OC is an infinitive, a base, or a participial verb form.

- *Finite:* The Queen of Spain learned that Columbus had reached the New World. (E)
- *Nonfinite:* Jake volunteered to work the difficult math problem on the blackboard. (N)

Complex sentences with relative clauses (RC): Relative clauses follow nouns and provide additional information about that noun. Typically they begin with a relative pronoun (*that*, *who*, *whose*, *which*). A relative clause can modify any noun in a sentence, whether the noun functions as the main clause subject or object. Four structural subtypes were trained in this study:

- *Subject, Subject:* The candidate that wins the primary advances to the main election. (E)
- *Subject, Object:* The used car that Jake bought had major problems from the beginning. (N)
- *Object, Subject:* The legislature passed a law that reduced car emissions by half. (E)
- *Object, Object:* We met the new band director that the school district hired over the summer. (N)

⁷N in parentheses refers to narrative content (about a common life experience). E in parentheses refers to expository (informational) content.