ASHA American Speech-Language-Hearing Association

JSLHR

Research Note

Sentence Completion Expressive Language Task–Preschool: Preliminary Evidence of Validity and Feasibility for the Neuroimaging Environment

Karla N. Washington,[a,b,c,d,e] Leslie E. Kokotek,[e,f] Mekibib Altaye,[g,h] Jennifer Vannest,[i,j] Mark DiFrancesco,[k,l] and Thomas Maloney[l]

[a]Department of Speech-Language Pathology, University of Toronto, Ontario, Canada [b]Rehabilitation Sciences Institute, University of Toronto, Ontario, Canada [c]The Institute for Education Research, University Health Network, Toronto, Ontario, Canada [d]Children's Voices Centre, Charles Sturt University, Bathurst, New South Wales, Australia [e]University of Cincinnati, OH [f]Division of Communicative Disorders, University of Louisville, KY [g]College of Medicine, University of Cincinnati, OH [h]Division of Biostatistics and Epidemiology, Cincinnati Children's Hospital, OH [i]Department of Communication Sciences and Disorders, University of Cincinnati, OH [j]Division of Speech-Language Pathology, Cincinnati Children's Hospital Medical Center, OH [k]Department of Radiology, Cincinnati Children's Hospital Medical Center, OH [l]Department of Radiology, University of Cincinnati, OH

ARTICLE INFO

Article History:
Received May 15, 2025
Revision received July 18, 2025
Accepted July 15, 2026

Editor-in-Chief: Julie A. Washington
Editor: Sara T. Kover

https://doi.org/10.1044/2026_JSLHR-25-00365

ABSTRACT

Purpose: This preliminary investigation occurred across two studies and examined validity, discriminant accuracy, and feasibility of the Sentence Completion Expressive Language Task–Preschool (SCELT-P) for characterizing language production in a neuroimaging environment for typically developing (TD) preschoolers and those with developmental language disorder (DLD).
Method: Two separate studies were completed. Study 1 (validity) included 82 preschoolers (50% female; 69 TD, 13 DLD) aged 4;0 (years;months) to 5;11 (M_{age} = 4;7). All children were assessed using standardized language measures (e.g., Clinical Evaluation of Language Fundamentals Preschool–Second Edition), a play-based language sample (analyzed using Developmental Sentence Scoring), and the 30-item SCELT-P (outside the neuroimaging environment), later streamlined to 20 items (SCELT-P Revised [SCELT-PR]). Study 2 (feasibility) included 14 preschoolers (M_{age} = 4;9, 43% female; nine TD, five with DLD) that were assessed using the SCELT-P (inside the neuroimaging environment) and functional magnetic resonance imaging using a sparse acquisition approach with a "listen" (passive) and "listen + speak" (active) condition implemented using DirectRT software.
Results: In Study 1, the SCELT-P and SCELT-PR demonstrated high reliability, including interrater (intraclass correlation coefficient = .99), test–retest (r = .86 and .80, respectively), and internal consistency (Cronbach's α > .86 and .83, respectively), and similar strong criterion validity (r = .53–.84). Area under the curve analyses (.91) revealed both versions also evidenced fair sensitivity (.88 and .87) and specificity (.85) rates with a cutoff score of 15.5 and 9.5, respectively. In Study 2, 89% of TD preschoolers and 60% of preschoolers with DLD successfully completed the SCELT-P/R inside the neuroimaging environment, with activation maps illustrating relatively similar neurobiological characteristics across preschoolers in some brain regions.
Conclusion: The SCELT-P, including its streamlined version, the SCELT-PR, holds promise as a reliable and valid task for including 4- to 5-year-olds in the neuroimaging environment to inform neural and cognitive systems involved in language development.
Supplemental Material: https://doi.org/10.23641/asha.34543209

Correspondence to Karla N. Washington: karla.washington@utoronto.ca.
Disclosure: *Karla N. Washington is lead developer and Jennifer Vannest is co-developer of the Sentence Completion Expressive Language Task–Preschool task. The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

Our ability to visualize the structure and function of the developing brain has experienced remarkable increases over the past 2 decades (Abbott & Love, 2023; Bishop, 2013; Ullman et al., 2024). Notably, measurement of assessment and intervention outcomes has been markedly

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

influenced by advances in neuroimaging, that is, brain imaging research (Chouinard et al., 2016; Hoyda et al., 2025; Ning et al., 2025). Neuroscientific methods such as functional magnetic resonance imaging (fMRI) are being included as part of assessment and intervention research to understand neurodevelopmental disorders such as developmental language disorder (DLD), a prevalent pediatric disorder. On average, two children in every U.S. classroom (approximately 7%) are diagnosed with DLD, a disorder often characterized by hallmark deficits in the production and perception of grammatical sentences (Bishop et al., 2017; Leonard, 2013). These children face social, economic, and academic challenges to the realization of their adult potential (Bishop, 1992; McGregor, 2020). While the behavioral profile of children with DLD is becoming better-characterized, the neurocognitive basis remains relatively unclear (Ullman et al., 2024).

Using neuroimaging tools to describe language learning offers an exciting opportunity to further basic knowledge of neural and cognitive systems in preschoolers with and without DLD. These technologies allow language researchers to evaluate, measure, and refine theories related to the neural substrates of both typical language processing (Hickok & Poeppel, 2007; Uddén et al., 2022) and language disorders (Badcock et al., 2012; Bishop, 2013; Ullman et al., 2020). This orientation in thinking was illustrated more than a decade earlier by Coltheart and McArthur (2012) who posited that if brain changes are occurring and they are correlated with behavioral changes, then the neuroimaging data could inform us about the biological underpinnings of change. However, for preschoolers with DLD, and especially those who demonstrate hallmark deficits in morphosyntax, our dimension of interest, no tools seem to exist that evaluate expressive grammar and demonstrate validity, reliability, and feasibility for use in the neuroimaging environment. This paucity in available tools limits our understanding of potential mechanisms underlying this prevalent disorder. Existing measures such as the Test for Reception of Grammar and the Structured Photographic Expressive Language Test–Third Edition (SPELT-3), among others, demonstrate validity for administration modalities outside the neuroimaging environment. Their administration relies on either extended clinician–child interaction and/or physical test materials, which are incompatible with the constraints of a neuroimaging setting. In response to this need, this research note explores the validity and feasibility of a new preschool language production task for use in the neuroimaging environment, the Sentence Completion Expressive Language Task–Preschool (Washington et al., 2013, 2018).[1] In these efforts, the authors offer a discussion on fMRI as a neuroimaging method and how it proffers our understanding of DLD.

Neuroimaging and DLD

Presumably, children with DLD have differences from typically developing (TD) children in the neural mechanisms that support language skills related to their language deficits. The extant neuroimaging literature suggests that, compared to controls, children with DLD or adults with a history of DLD demonstrate structural or functional differences in regions known to underlie domain-specific language skills (i.e., phonological, semantic, syntactic, auditory processing) in typical adults (e.g., Abbott & Love, 2023; J. C. Lee, Dick, & Tomblin, 2020; Verly et al., 2019; Vydrova et al., 2015). Specifically, in several studies of individuals with DLD, a left-lateralized frontotemporal language network (including Broca's area) has been shown to be affected in terms of both structure (gray matter morphology) and function (see Mayes et al., 2015, for a review; Pigdon et al., 2020). However, DLD may also be associated with structural or functional abnormalities in brain regions known to support more domain-general processes, such as the basal ganglia, motor/supplementary motor regions, and the cerebellum (see also Krishnan et al., 2022; also reviewed in Mayes et al., 2015). In one fMRI study using a connectivity approach (Plante et al., 2017), no differences between TD adults and adults with a history of DLD were observed. In contrast, in studies using diffusion imaging ("structural connectivity"), white matter tracts connecting these regions have been shown to be broadly affected (J. C. Lee, Dick, & Tomblin, 2020; M. H. Lee, O'Hara, et al., 2020; Verly et al., 2019; Vydrova et al., 2015).

These existing studies have been somewhat limited in their ability to draw conclusions about the contributions of specific brain networks in DLD. While differences between DLD and typical groups were documented in several studies, the direction of effects (hypo- or hyperactivation in fMRI, increases or decreases in diffusion measures) varies. This variability may relate to patterns of functional activation and connectivity elicited by task-based fMRI, which may be highly affected by task and task performance. For example, when task performance was controlled for, a recent large-scale study by Krishnan et al. (2022) did not find expected differences in activation between 10- and 15-year-old children with DLD and TD peers during a verb generation task. They noted that tasks engaging more complex language constructs, as well as connectivity analyses, may be more sensitive to differences in function. In addition, Krishnan et al. suggest these differences may also be more detectable in younger children who are more actively refining their language system. The findings from Krishnan et al. may also indicate that isolated verb generation may not directly reflect

[1]Citation year 2013 reflects when the ISBN was obtained for publication purposes, which also coincided with the year the SCELT-P was first used. Citation year 2018 reflects when the ISBN for publication purposes was obtained, not the year the task (SCELT-P Revised) was first used, which was in 2014.

the core morphosyntactic deficits that are prototypical of DLD, and therefore, their null findings could be reflective of the language domain targeted rather than a true absence of neural differences. As such, this quandry underscores the need for developmentally appropriate imaging tools that are designed to elicit expressive grammar beyond the single-word level.

A possible framework for helping to understand these differences is the procedural circuit deficit hypothesis (PDH; Ullman et al., 2020). Described simply, the PDH suggests that the memory–language connection is at the core of language learning difficulties for children with DLD. More specifically, the PDH underscores that children with DLD are prone to experiencing disruptions in procedural memory, which is a foundational skill necessary for rule-based learning (such as grammatical processing). Importantly, the PDH considers not just structural abnormalities but also differences in neuroconnectivity. For instance, neuroimaging studies (see Ullman et al., 2024, for a review) that align with the PDH have demonstrated underactivation and disrupted connectivity in procedural memory networks in those with DLD, especially between the caudate nucleus and frontal brain regions necessary for procedural learning.

SCELT-P

Due to task requirements (e.g., verbally responding while in the scanner) and neuroimaging procedures (e.g., remaining still), most neuroimaging research on language skills has been conducted with children aged 6 years and older or with adults (Bishop, 2013). Furthermore, most established tasks for children, such as a verb generation task, investigate language skills at the single-word level only (e.g., Holland et al., 2007; Krishnan et al., 2022). To address this gap, prior researchers have developed highly effective techniques (e.g., use of a block-periodic design where tasks of interest were alternated with a control task, implementing fMRI tasks that have clinical relevance and are also compatible with the neuroimaging environment), suitable for use while scanning 4- to 5-year-olds (Byars et al., 2002; Holland et al., 2007), with at least a 75% success rate in children as young as 4 years of age (Vannest et al., 2014). These developments informed the design and implementation of a computer-based task for fMRI, the SCELT-P (Washington et al., 2013).

The SCELT-P was designed to explore grammatical functioning beyond just the single-word level in 4- to 5-year-old preschoolers who are TD and those with DLD, particularly when their DLD is characterized by hallmark morphosyntactic deficits. This is an important consideration in measuring children's language skills (Sealey & Gilmore, 2008; Washington & Warr-Leepr, 2013; Yarian et al., 2021) as previous verb generation tasks are often limited to generating words in isolation. Language production tasks such as the SCELT-P are desirable for fMRI because understanding language production mechanisms is crucial for characterizing grammatical deficits and potential changes with treatment in preschoolers with DLD that manifests in impaired morphosyntax (Washington et al., 2014, 2015). Not designed to diagnose language disability, the SCELT-P was informed by Brown's grammatical morphemes (Brown, 1968, 1973) and information on grammatical development (e.g., Hadley, 2014; Marchman et al., 1999; Miller, 1996; Paul & Alforde, 1993; Paul & Norbury, 2012; Valian, 2006). In this vein, early developing items such as present progressive tense (e.g., Valian, 1992) were listed ahead of later developing items such as reflexive pronouns. Additionally, scale development and evaluation included a binary accuracy scale reflecting "1" (*correct*) and "0" (*incorrect*), which complemented other measures of child development including vocabulary (Expressive Vocabulary Test–Second Edition; Williams, 2007), nonverbal intelligence (Primary Test of Nonverbal Intelligence; Ehrler & McGhee, 2008), and our construct of interest, morphosyntax (SPELT-3 and Clinical Evaluation of Language Fundamentals Preschool–Second Edition [CELF-P2]).

Important to note is that the SCELT-P was initially developed as a 30-item measure containing 30 listen–speak pairs of copyright-free cartoon-like images presented in a somewhat developmental sequence. The SCELT-P allowed children to have at least one opportunity to produce each targeted morphosyntatic form. For assessment completion, children respond to the SCELT-P using a cloze procedure, requiring that they complete a simple sentence in a syntactic frame, for example, "Here a girl is sleeping" and "Here a girl is _____" (present progressive tense expected). To optimize future administrations, a streamlined version was developed that removed duplicate items with the lower accuracy on average across all participants. This streamlining resulted in 20 items containing 20 listen–speak pairs being included on the SCELT-P Revised (SCELT-PR; this streamlined process is returned to in the Method section). See Table 1 for SCELT-P (Washington et al., 2013) and SCELT-PR (Washington et al., 2018) for morphosyntactic structures.

To assess the utility of the SCELT-P and SCELT-PR, reliability estimates for performance-level scores are needed and should include internal consistency, interrater, and test–retest reliability. By definition, internal consistency reliability describes the extent to which responses on items on a scale relate to each other (Portney & Watkins, 2009). The Cronbach's alpha statistic is typically used to measure internal consistency and is deemed "acceptable" at ≥ .6 and "good" at ≥ .7. Interrater reliability, which refers to the degree of agreement in performance scoring between two independent raters, should demonstrate minimum reliability coefficients of .70 to indicate clinical utility (Nunnally &

Table 1. Sentence Completion Expressive Language Task–Preschool (SCELT-P) and SCELT-P Revised morphosyntactic structures.

Stimulus no.	Grammatical category	Accuracy	Stimulus no.	Grammatical category	Accuracy
1	Present progressive	84%	16	Third-person singular	48%
2	Present progressive[a]	90%	17	Third-person present tense irregular[a]	46%
3	Preposition	93%	18	Uncontactable auxiliary[a]	59%
4.	Preposition[a]	96%	19	Contractable copula[a]	51%
5	Regular plural[a]	89%	20	Contractable auxiliary[a]	46%
6	Irregular plural[a]	94%	21	Negation[a]	35%
7	Irregular past tense[a]	43%	22	Future tense	28%
8	Irregular past tense	39%	23	Future tense[a]	33%
9	Possessive noun[a]	57%	24	Subjective pronoun	59%
10	Uncontactable copula[a]	65%	25	Subjective pronoun[a]	80%
11	Article	83%	26	Possessive pronoun	72%
12	Article[a]	84%	27	Possessive pronoun[a]	80%
13	Regular past tense[a]	59%	28	Objective pronoun	80%
14	Regular past tense	45%	29	Objective pronoun[a]	83%
15	Third-person singular[a]	48%	30	Reflexive pronoun[a]	6%

[a]Items from the SCELT-P identified for use in the SCELT-P Revised.

Bernstein, 1994). Test–retest reliability, on the other hand, refers to the degree of agreement in performance scoring by the same rater over a specified time frame.

In addition to being reliable, a tool should also demonstrate evidence of validity (Gay, 1985; Messick, 1993). One measure of validity is criterion validation. Criterion validation is an important form of validity evidence for determining psychometric properties for tools to establish the direction and magnitude of the relationship between the target tool and other standard tools that measure similar abilities (Gay, 1985; Messick, 1993), with Cohen (1988) describing correlation magnitudes as minimal ($.1 < |r| < .3$), moderate ($.3 < |r| < .5$), and strong ($|r| > .5$). To establish its psychometric properties, the SCELT-P and SCELT-PR scores were compared to scores on established well-known tools measuring morphosyntax, specifically, the CELF-P2 (Wiig et al., 2004) Word Structure (WS; 24 items) subtest, the SPELT-3 (53 items; Dawson et al., 2003), and a play-based language sample analyzed using the Developmental Sentence Scoring (DSS; L. Lee, 1974; L. L. Lee & Canter, 1971) metric. These measures are described briefly below.

CELF-P2

The CELF-P2 is a comprehensive language assessment used to evaluate language abilities and to identify language deficits in children aged 3;0 (years;months) to 6;11. The Core Language Score (WS, Sentence Structure [SS], and Expressive Vocabulary subtests), which has > 80% sensitivity and specificity rates, has been used in the classification of DLD. The WS subtest evaluates expressive morphology by examining a child's ability to produce grammatical forms (e.g., plurals) in sentence completion tasks. Performance on this subtest can be used to inform hallmark deficits in morphosyntax. When determining the presence of DLD, clinicians can compare a child's standard score to normative data and determine whether they fall below 1.5 *SD*s of the average expected performance for their age level.

DSS

The DSS metric is a form of language sample analysis used to evaluate the grammatical complexity of spontaneous language productions in young children ages 2;0–6;11 (L. Lee, 1974; L. L. Lee & Canter, 1971). The DSS is calculated by assigning weight points to selected grammatical forms, such as verb and syntactic structures across eight grammatical categories/subscales, and calculating a total score. Historically, scores falling below the 10th percentile are indicative of potential language disorder, including DLD (cf. Leonard & Deevy, 2020; Washington et al., 2011; Washington & Warr-Leepr, 2013). Recent evidence shows that total scores and the main verb subscale offer the most sensitivity in informing typical versus atypical grammatical profiles (Bernstein Ratner et al., 2026).

SPELT-3

The SPELT-3 is a standardized language assessment that evaluates expressive morphosyntax in children aged 4;0–9;11. It identifies language disorder (e.g., DLD) by comparing a child's performance to normative expectations, with lower scores indicating difficulties in the production of age-appropriate grammatical structures. The SPELT-3 manual indicates, as part of its analysis and interpretation of scores, a standard score of 100 is average, with an *SD* of 15. Follow-up research using the SPELT-3 has supported the use of a standard score of 95 as the optimal cutoff for achieving > 90% sensitivity and specificity in classification as

DLD (cf. Perona et al., 2005). This cut-point was considered in concert with scores on another language measure (the Test for Examining Expressive Morphology; Shipley et al., 1983) and informal observation to inform typical development or DLD.

The Current Study

The purpose of this preliminary investigation was twofold. First (Study 1), to establish the initial psychometric properties of the SCELT-P and SCELT-PR by providing evidence of reliability, criterion, and construct validation. Relatedly, the authors were also interested in exploring discriminant accuracy. Second (Study 2), we sought to assess feasibility in the neuroimaging environment by way of success/completion rates, with a sample of preschoolers with and without DLD whose primary area of impairment reflected morphosyntactic errors.[2] In this study, we also explored what insights could be learned about the neurobiological underpinnings of grammatical abilities in preschoolers with and without DLD. By providing early evidence of psychometric properties, namely, reliability and validity (Study 1), along with utility in a neuroimaging environment (Study 2), this research aims to enhance our understanding of language development in preschoolers using tools with psychometric evidence. By focusing specifically on 4- to 5-year-old children, we seek to offer novel clinical insights into language development profiles (specifically morphosyntax) in this age group, which are often underrepresented in neuroimaging research (cf. Ullman et al., 2020, 2024).

Study 1

Method

Ethical Approval

Ethical approval was obtained from the University of Cincinnati's institutional review board (IRB; No. 2013-5890) in tandem with a reliance IRB at Cincinnati Children's Hospital Medical Center (No. 2021-0716) prior to the commencement of this research. The University of Toronto's research ethics board (No. 00043895) also provided approval related to the use of previously collected data. Parents provided consent for their children to participate. Children also engaged in verbal assent procedures.

[2]McGregor et al. (2025) make note that DLD is a multidimensional spectral system that includes, but is not limited to, phonology, morphosyntax, semantic representations, figurative language, and pragmatics. For the sample of preschoolers with DLD in the current investigation (Study 1 and Study 2), we are referring to children experiencing primary deficits within the dimension of morphosyntax.

Participants

Preschoolers were recruited from community settings, including early learning centers, clinics, hospitals, and preschools in Cincinnati, Ohio, between 2013 and 2019. Study 1 included 82 preschoolers aged 4;0–5;11 (M_{age} = 4;7, SD = 0.47), with 41 girls and 41 boys. Based on intake testing, discussed below, children were categorized as either TD (n = 69) or having DLD (n = 13). Using a multimeasure converging evidence approach, preschoolers categorized as TD demonstrated average to above-average performance on all standardized tasks and the language sampling task and met screening inclusion for the study (i.e., typical hearing, cognition, and oral-motor skills). Specific assessments for group classification are described in more detail below.

Since our dimension of interest was morphosyntax, these assessments included the CELF-P2 WS subtest as a primary measure, the SPELT-3 as a descriptive measure of morphosyntactic abilities, and a play-based language sample scored using the DSS metric. Preschoolers categorized as having DLD also met inclusion criteria but demonstrated below-average performance on our primary measure of expressive grammar (i.e., $\geq$ 1.5 SDs below the mean on the CELF-P2 WS subtest). They also demonstrated below-average performance and the SPELT-3 and on DSS scoring. Thresholds of below-average performance were informed by test manuals as well as guidelines offered in prior published works including our own (e.g., DSS, $\leq$ 10th percentile [Leonard et al., 2006; Leonard & Deevy, 2020; Washington et al., 2011; Washington & Warr-Leepr, 2013]; SPELT-3 cut-point 95 [Perona et al., 2005]). These data were considered holistically to inform expressive grammar profiles. Preschoolers' receptive language at the word and sentence levels (Peabody Picture Vocabulary Test–Fourth Edition [PPVT-4; Dunn & Dunn, 2007]; CELF-P SS [Wiig et al., 2004]) and nonverbal IQ (Kaufman Brief Intelligence Test–Second Edition [KBIT-2; Kaufman & Kaufman, 2004]) also provided descriptive information about preschoolers' language skills.

These criteria for identifying and describing language impairment aligned with earlier and recent discussions regarding DLD that described processes for using a combination of formal and informal assessment methods (e.g., Finneran et al., 2009; Leonard & Deevy, 2020; Perona et al., 2005; Washington et al., 2011; Westby, 2007) to infuse ecological validity into our diagnostic efforts. Importantly, we note that the unbalanced group sizes is reflective of the broader occurrence of DLD (i.e., ~7%) within the population. That said, this ratio was also reflective of the difficulty in recruiting a larger sample of children with DLD.

Instruments

SCELT-P. Field testing (45 graduate communication sciences and disorders [CSD] students, 25 CSD faculty, 10

preschoolers [four with DLD and six TD], and five non-CSD adults) revealed that 99% of respondents described the SCELT-P images as child friendly and that the task could be completed on a computer. Although not intended to diagnose language disorder, adults who responded also stated that "morphosyntactic/grammar skills were being evaluated," providing some evidence of face and content validity (Washington et al., 2014) for the SCELT-P (Washington et al., 2013).

Participants completed the computer-based SCELT-P with a clinician as part of a battery of assessment tasks (outside the neuroimaging environment). Its use of a standard testing paradigm (e.g., a cloze procedure) and a computer-based platform provided ecological validity in testing format. Designed and presented using DirectRT software, https://www.empirisoft.com, the SCELT-P contains pictures and audio recordings to facilitate its completion. The initial SCELT-P included 30 items (30 pairs). These 30 pairs of images were presented in a fairly developmental sequence using different stimuli: (a) stimulus (control) – listening only (n = 15) and (b) response – listening and speaking (n = 15). Practice items were included to establish the expected routine and to facilitate accuracy in responses; the question "Are you ready?" was asked prior to presenting each stimulus item. Children responded to the statements from the SCELT-P program using a cloze procedure within a specific syntactic frame (see Figure 1). A streamlined version with 20 items, the SCELT-PR, was later developed based on item accuracy across morphosyntactic opportunities presented per target feature. For example, the SCELT-P included two opportunities to produce the present progressive tense (84% first opportunity, 90% second opportunity) and preposition (96% first opportunity, 93% second opportunity). The item with the higher accuracy on average across all preschoolers was included in the streamlined version. Both versions were compared to well-known measures of expressive grammar to establish psychometric merits (see Supplemental Material S1 for details).

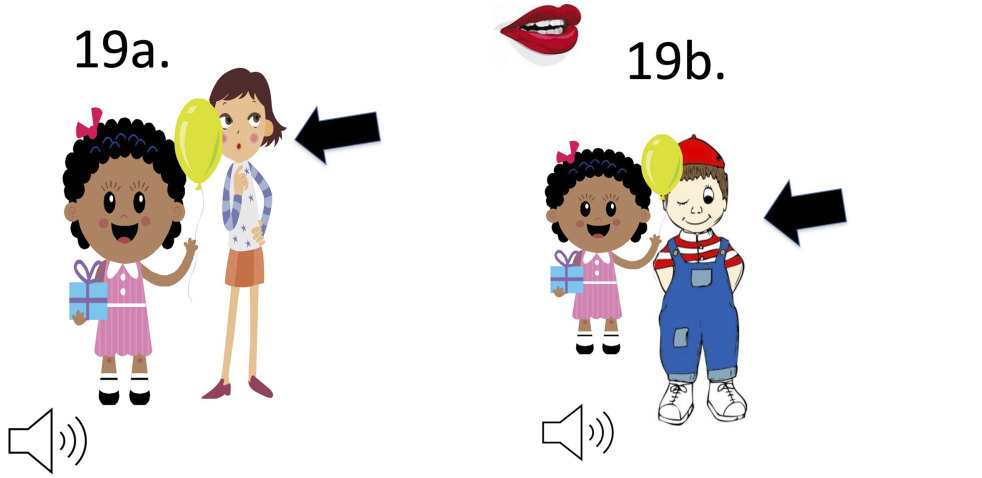
Measures of morphosyntax. To establish the validity of the SCELT-P and SCELT-PR, expressive grammar performance was also measured for each preschooler using well-known assessments in speech-language pathology that emphasized morphosyntax. These included the CELF-P2 WS subtest (24 items; Wiig et al., 2004), the SPELT-3 (53 items; Dawson et al., 2003), and a spontaneous language sample scored using DSS (eight grammatical categories; L. Lee, 1974). Guidelines for scoring were adhered to for each measure. For the spontaneous language sample, 15-min play-based samples were collected. The same doll house, household items, and people, along with books, were used for each language sample session.

Procedure

Intake Testing

Behavioral activities were completed at children's early learning center, preschool, or clinic setting (i.e., where the participant was recruited). These activities were divided

Figure 1. Example of stimulus items for the SCELT-P and SCELT-P Revised. The task represents a cloze procedure using a syntactic frame. SCELT-P = Sentence Completion Expressive Language Task–Preschool.



into screening (5- to 7-min) and assessment (60- to 75-min) sessions. Graduate students supervised by a certified speech-language pathologist (SLP) completed all evaluations. All preschoolers demonstrated (a) binaural hearing at 20 dB for 1, 2, and 4 kHz; (b) parent-reported absence of neurological deficits or pervasive developmental disorders; (c) age-based criterion on the Oral Motor subtest of the Diagnostic Evaluation of Articulation and Phonology (DEAP) (Dodd et al., 2006); and (d) standard scores of ≥ 85 (M = 105.90, SD = 11.23, range: 85–135) on the KBIT-2 (Kaufman & Kaufman, 2004). Importantly, while a very conservative approach, average performance on the KBIT-2 was included to help rule out intellectual disability. Based on the preschoolers' performance, no participants from either the TD or DLD group were excluded on the basis of KBIT-2 standard scores.

As mentioned above, for assessment of TD and of DLD, preschoolers completed receptive language and expressive grammar measures. These assessments were administered in a counterbalanced fashion to mitigate a potential order effect. Receptive language measures included the PPVT-4 (Dunn & Dunn, 2007) and the SS subtest of the CELF-P2 (Wiig et al., 2004). Expressive grammar measures included the CELF-P2 WS subtest, the SPELT-3, and the DSS-scored spontaneous language sample. Expressive grammar test administration also included the SCELT-P, but this assessment was not used as a diagnostic tool. Important to note is that n = 13 (18.8%) of preschoolers classified as TD in Study 1 did not meet the optimal cut-point of 95 on the SPELT-3 (M = 91.23, SD = 3.83, range: 86–95) as noted in Perona et al. (2005). However, these preschoolers, who scored well above the −1.5-SD cut-point on the CELF-P2 WS (M = 100.76, SD = 9.09, range: 90–120), used expected morphemes > 90% of the time based on SLP informal observation during play and scored in the average range (> 25th percentile) on the DSS. Similar to the Perona et al. (2005, p. 105), a holistic approach using both formal and informal measures was used to reconcile these preschoolers' profile as TD and not as DLD (see Table 2 for participant demographics and test performance).

Data Analysis

A cross-sectional design was used to gather evidence about the reliability and validity of the SCELT-P and SCELT-PR. Evidence regarding the discriminant accuracy, defined as good (90%), fair (80%–89%), or unacceptable (< 80%), was also obtained (Plante & Vance, 1994). All data were analyzed using SPSS Version 20.0.0.

For Study 1, reliability analyses included three types of evidence. First, internal consistency for both SCELT-P versions was calculated using Cronbach's alpha. This process included evaluating the extent to which individual items on the SCELT-P were correlated with each other and the degree to which they measured the same construct. Second, interrater reliability to determine the degree of agreement between different raters was evaluated using intraclass correlation coefficient (ICC) to examine the correspondence between the SCELT-P scores assigned by different raters. Third, test–retest reliability was conducted using a 1-week interval to evaluate consistency across 16 preschoolers.

Validity evidence of the SCELT-P and SCELT-PR was explored using criterion and content validation (Streiner & Norman, 2008; Streiner et al., 2014). For criterion validation, this process included conducting Pearson correlations to examine the relationship between SCELT-P and SCELT-PR scores (out of 30 and out of 20, respectively) and scores on other well-validated measures of expressive grammar that were included as part of the intake testing, namely, the CELF-P2 WS subtest, the SPELT-3, and DSS analysis of play-based language samples. Additionally, as mentioned above, field testing with relevant stakeholders that included 25 CSD professionals (among others) revealed 99% agreement that the SCELT-P was child friendly and appropriate for computer-based completion (i.e., face validity) and measured morphosyntax (i.e., content validity). These findings offer early, but robust, evidence and suggest that the SCELT-P measures its intended construct (i.e., the morphosyntax of preschool children).

With these data, the discriminant accuracy of the SCELT-P and SCELT-PR was also established using logistic regression as a discriminant analysis for a first step in subsequently obtaining a receiver operating characteristic (ROC) curve and area under the curve (AUC) to find the optimal cut-point. Although this study primarily included monolingual English-speaking children, this approach is similar to that in the published literature when classifying TD from disorder across monolingual and multilingual preschoolers (e.g., Ebert et al., 2020; Justice et al., 2019; Leon et al., 2021; Pace et al., 2022).

Results

Reliability

The SCELT-P tasks demonstrated "good" internal consistency as measured by Cronbach's alpha (SCELT-P, α = .86; SCELT-PR, α = .83). Interrater reliability using a two-way mixed ICC was found to be very high, with an interrater agreement of .99 (95% CI [.996, .999], p < .001) for the SCELT-P and .99 (95% CI [.991, .999], p < .001) for the SCELT-PR. The tasks also exhibited a high test–retest reliability for total scores (SCELT-P, r = .86, p < .001; SCELT-PR, r = .80, p < .001).

Validity

Evidence of criterion validation was noted by way of examining relationships between the SCELT-P/SCELT-PR

Table 2. Participant demographics and assessment scores for Study 1.

Child and assessment details[a,b]	Gender	Age (months)	KBIT-2	PPVT-4	CELF-P2 (SS)	CELF-P2 (WS)[c]	SPELT-3	DSS spontaneous language sample	SCELT-P (30 items)	SCELT-PR (20 items)
TD (*n* = 69)	*n* = 35 boys *n* = 34 girls	*M* = 56.21 *SD* = 5.51 Range: 48–68	*M* = 106.86 *SD* = 11.28 Range: 85–135	*M* = 110.29 *SD* = 13.93 Range: 84–142	*M* = 107.54 *SD* = 13.47 Range: 80–140	*M* = 107.54 *SD* = 12.99 Range: 90–145	*M* = 102.80 *SD* = 12.77 Range: 86–128[d]	*M* = 7.14 *SD* = 1.08 Range: 5.24–10.46	*M* = 20.31 *SD* = 4.28 Range: 9–29	*M* = 13.21 *SD* = 3.29 Range: 5–19
DLD (*n* = 13)	*n* = 6 boys *n* = 7 girls	*M* = 57.92 *SD* = 6.57 Range: 49–73	*M* = 100.85 *SD* = 9.86 Range: 85–118	*M* = 88.46 *SD* = 13.72 Range: 73–118	*M* = 88.85 *SD* = 12.10 Range: 65–105	*M* = 73.46[e] *SD* = 5.55 Range: 60–80	*M* = 73.54[e] *SD* = 5.23 Range: 63–78	*M* = 5.52[e] *SD* = 0.77 Range: 4.10–6.93	*M* = 10.46 *SD* = 5.65 Range: 4–20	*M* = 6.15 *SD* = 3.8 Range: 2–13

Note. KBIT-2 = Kaufman Brief Intelligence Test–Second Edition; PPVT-4 = Peabody Picture Vocabulary Test–Fourth Edition; CELF-P2 = Clinical Evaluation of Language Fundamentals Preschool–Second Edition; SS = Sentence Structure; WS = Word Structure; SPELT-3 = Structured Photographic Expressive Language Test–Third Edition; DSS = Developmental Sentence Scoring; SCELT-P = Sentence Completion Expressive Language Task–Preschool; SCELT-PR = Sentence Completion Expressive Language Task–Preschool Revised; TD = typically developing; DLD = developmental language disorder.

[a]*n* = 31 African American children, *n* = 44 Caucasian children, *n* = 2 Latine children, and *n* = 5 multiracial children. [b]All children passed the Oral Motor subtest of the Diagnostic Evaluation of Articulation and Phonology and a binaural hearing screening at 25 dB at 1, 2, and 4 kHz. [c]Primary measure for group classification at less than –1.5 *SD*s below the mean. [d]Of the participants, 18.8% (*M* = 91.23%, *SD* = 3.83) did not meet the cut score of 95 based on Perona et al. (2005) but were within the average range per SCELT-3 manual (*M* = 100, *SD* = 15) and the CELF-P WS (*M* = 100.77, *SD* = 9.09) and in a spontaneous play-based context with a speech-language pathologist used expected morphemes that > 90% of the time. These preschoolers also scored well above the 10th percentile cut based on their DSS scores. [e]≤ 10th percentile.

and the other well-validated measures previously described. A strong positive correlation was observed between the CELF-P2 WS subtest and the SCELT-P measures ($r = .84$, $p < .001$). A strong positive relationship was also found between the SPELT-3 and SCELT-P ($r = .73$, $p < .001$) and SCELT-PR ($r = .72$, $p < .001$), while a moderate positive correlation was observed between the SCELT-P measures ($r = .55$, $p < .001$) and DSS scores ($r = .53$, $p < .001$).

Discriminant accuracy. To address discriminant accuracy, we completed an ROC curve and AUC analyses to demonstrate the strength of task differentiation between typical development and DLD. A principal component analysis (PCA) was employed for feature extraction to begin the process for determining the discriminant accuracy of the SCELT-P and SCELT-PR. PCA of the SCELT-P indicated that 15 components accounted for 80.1% of the variance in scores, while 12 components accounted for 82.8% of the variance in scores in the revised version. A scree plot of the eigenvalues for the SCELT-P (see Figure 2A) revealed a distinct inflection point (i.e., bend) at the second principal component, suggesting that subsequent components contribute minimally to the overall variance in scores. A scree plot of the eigenvalues for the SCELT-PR (see Figure 2B) revealed a distinct inflection point (i.e., bend) after the first principal component, suggesting that subsequent components contribute minimally to the overall variance in scores.

For the SCELT-P, we fit a logistic regression to plot an ROC curve and AUC analyses (.91), further demonstrating that this task possesses strong differentiation, with fair sensitivity (.88) and specificity (.85) values supporting its utility in helping to distinguish typical development from DLD. A cutoff score of 15.5 ($SD = 1.13$ below TD mean) was identified as an optimal threshold for making this determination to find the true positive versus false positive. For the SCELT-PR, an ROC curve was also plotted, and AUC analyses (.91) further demonstrated that this version also possesses strong differentiation, with fair sensitivity (.87) and specificity (.85) values supporting its utility in helping to distinguish typical development from DLD. A cutoff score of 9.5 ($SD = 1.13$ below TD mean) was identified as an optimal threshold for making this determination.

Study 2

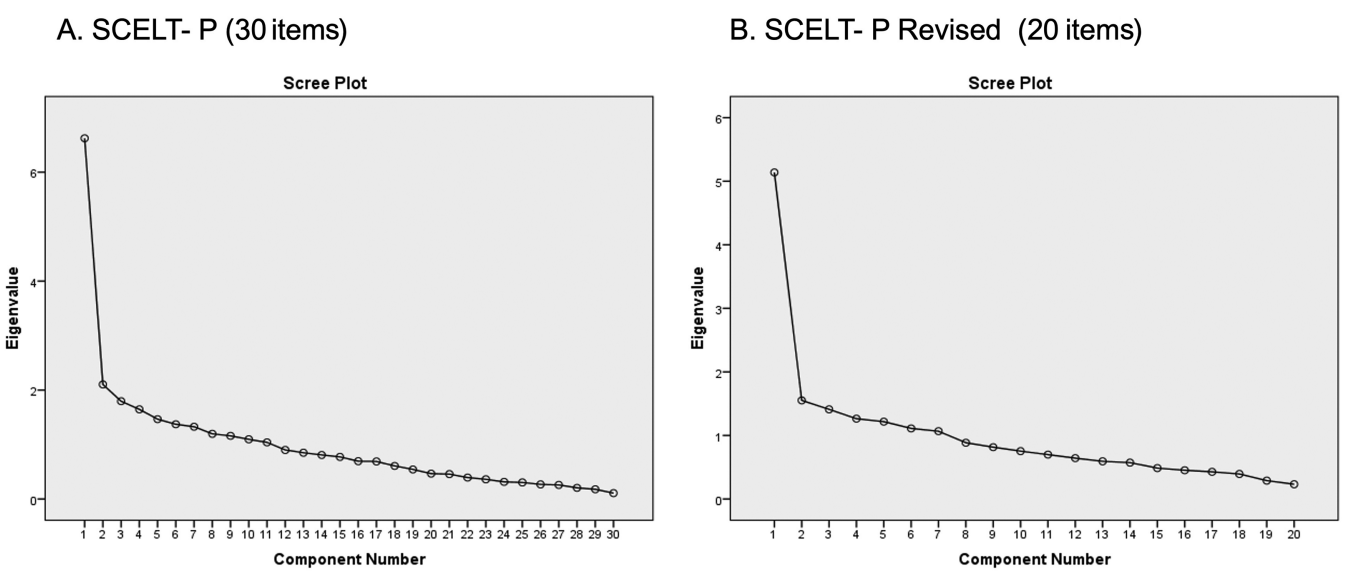
Method

The same procedures and ethical approval described in Study 1 were applied for Study 2. Parents and children provided consent and verbal assent, respectively. Preschoolers in Study 2 also assented to participating in the fMRI with their parents' consent. Importantly, preschoolers in this study did not also participate in Study 1.

Participants

Study 2 included 14 preschoolers (nine TD, five with DLD), aged 4;0–5;11 (M_{age} = 4;9) with $n = 6$ girls and $n = 8$ boys, who met the same inclusion criteria as noted in Study 1: (a) binaural hearing at 20 dB for 1, 2, and 4 kHz; (b) parent-reported absence of neurological deficits or pervasive developmental disorders; (c) age-based criterion on the Oral Motor subtest of the DEAP (Dodd et al., 2006);

Figure 2. SCELT-P and SCELT-P Revised discriminant accuracy. (A) Scree plot for SCELT-P. (B) Scree plot for SCELT-P Revised. SCELT-P = Sentence Completion Expressive Language Task–Preschool.



and (d) standard scores of ≥ 85 (Study 2 subset; M = 104.07, SD = 13.84, range: 85–138) on the KBIT-2 (Kaufman & Kaufman, 2004). The same classification process described in Study 1 was also used in Study 2 to categorize preschoolers as TD or having DLD. Please see Table 3 for demographic and assessment performance details.

Neuroimaging Environment

Neuroimaging was conducted using a Philips 3-T Achieva MRI system with high-performance Omega HP gradients (45 mT/m, 200-mT/m/ms slew rate), a 108-channel digital RG chain (Philips Medical Systems), and a 32-channel head coil. Avotec audiovisual systems were used including a Silent Vision 6011 projection screen and a head-coil mounted mirror for fMRI stimuli presentation. For the SCELT-P production task, preschoolers' responses were audio-recorded using an MRI-compatible microphone in concert with the DirectRT program.

Procedure

In addition to behavioral testing as outlined in Study 1, preschoolers in Study 2 also participated in neuroimaging sessions. Importantly, the neuroimaging paradigm was successfully tested on an adult participant per Cincinnati Children's Hospital Medical Center (CCHMC) policy prior to its use with pediatric participants. See Figure 3 for adult performance on the SCELT-P. All imaging sessions were completed at CCHMC, and preschoolers' parents or designee accompanied them to the session. Each preschooler completed an approximately 30-min neuroimaging session. Prior to the imaging session, families were provided with training resources (e.g., video link, materials) to orient them to the neuroimaging environment. For feasibility, 11 preschoolers completed the SCELT-P, and following the streamlining process, an additional three children completed the SCELT-PR. Given the mostly identical correlations between both versions as mentioned above, completion of either assessment was considered valid. However, to ensure visualization of activation reflected the same functional tasks, only the 20 items that overlapped between the two versions were used for visual analyses. Likewise, only the completion of these 20 items was used in determining preschoolers' success rate in the neuroimaging environment.

Certified imaging technologists, along with members of the research team, conducted the imaging procedures and addressed participant queries. On the scanning day, participants received an in-person neuroimaging suite orientation to reduce anxiety and underwent metal screening for safety. Participant orientation included extra time (up to 30 min) for the participant to get comfortable and explore the scanner environment in the MRI scanner room. A research team member and magnetic resonance technologist worked with the participant to demonstrate how the child would be positioned in the scanner as well as how they could hear audio and see video stimuli, with reminders to remain still during scanning. Cushions were positioned around the participant and in the head coil as needed for comfort and to discourage motion. Once positioned comfortably in the scanner, the child could view a child-friendly movie of their choice and typically did not need a team member or parent in the room during scanning. In some cases, a parent or study team member remained in the scanner room with the participant for the duration of the scan, at participant or parent request, but this was generally discouraged to minimize head motion associated with looking at/talking to an individual in the room. The technologist was able to maintain contact with participants using the speaker system throughout the session to ensure task compliance and safety, as well as visual observation from the adjacent scanner control room window.

Each session included high-resolution T1-weighted structural imaging and the SCELT-P during fMRI. Trials involved paired conditions: a baseline "listen" (cue sentence and picture) and an active "listen + speak" (partial sentence with cue to speak). To mitigate task novelty and difficulty, all preschoolers practiced unique SCELT-P tasks outside the scanner with nonexperimental stimuli (novelty) and completed in the same sequence (difficulty). This meant that children completed practice items to ensure they had the opportunity to learn the testing format prior to entering the scanner. However, to ensure item learning did not occur prior to entering the scanner, children completed unique practice items that were not also included on the SCELT-P or SCELT-PR. This process allowed children to learn the test procedure while maintaining novelty once inside the scanner. Once in the scanner, children's productions were audio-recorded for later analysis. During the structural MRI, preschoolers watched a child-friendly movie of their choice (see Table 4 for imaging parameters). fMRI was acquired using a sparse acquisition approach allowing for scanner quiescence during the presentation of the auditory stimuli and recording of the child's verbal responses (Schmithorst & Holland, 2004). These auditory stimuli and responses took place during 8-s quiet intervals, followed by 6 s of fMRI acquisition regardless of which SCELT-P version was used (see Table 4).

Activation Maps

For Study 2, the percentage of TD preschoolers and preschoolers with DLD who successfully completed each structural (no excessive motion or artifacts) and functional imaging (remained in the scanner for the full duration of the SCELT-P/R assessment and produced a corresponding audio file and imaging data) task is reported in the Results

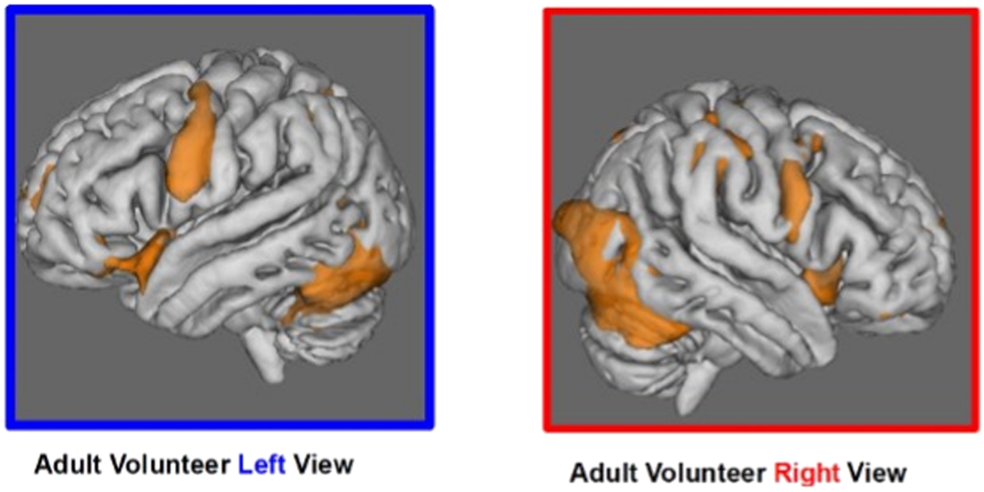
Table 3. Participant demographics and assessment scores for Study 2.

Child and assessment details[a,b]	Gender	Age (months)	KBIT-2	PPVT-4	CELF-P2 (SS)	CELF-P2 (WS)	SPELT-3	DSS spontaneous language sample	SCELT-P (30 items)	SCELT-PR (20 items)
TD (n = 9)	n = 5 boys n = 4 girls	M = 60 SD = 4.75 Range: 52–70	M = 108.33 SD = 13.37 Range: 89–138	M = 107.56 SD = 11.28 Range: 88–120	M = 104.44 SD = 12.36 Range: 85–120	M = 102.22 SD = 10.34 Range: 85–120	M = 103.89 SD = 5.34 Range: 96–114	M = 7.45 SD = 1.05 Range: 5.82–8.98	M = 23.83 SD = 2.22 Range: 22–28	M = 11.7 SD = 5.67 Range: 1–18
DLD (n = 5)	n = 3 boys n = 2 girls	M = 61 SD = 6.04 Range: 53–66	M = 96.4 SD = 11.50 Range: 85–112	M = 88.46 SD = 14.35 Range: 69–104	M = 81.00 SD = 21.62 Range: 55–105	M = 68.00[c] SD = 9.75 Range: 55–75	M = 74.00[c] SD = 8.43 Range: 63–81	M = 5.24[c] SD = .76 Range: 4.54–6.44	M = 7.6 SD = 5.77 Range: 0–16	M = 7 SD = 4.64 Range: 4–15

Note. KBIT-2 = Kaufman Brief Intelligence Test–Second Edition; PPVT-4 = Peabody Picture Vocabulary Test–Fourth Edition; CELF-P2 = Clinical Evaluation of Language Fundamentals Preschool–Second Edition; SS = Sentence Structure; WS = Word Structure; SPELT-3 = Structured Photographic Expressive Language Test–Third Edition; DSS = Developmental Sentence Scoring; SCELT-P = Sentence Completion Expressive Language Task–Preschool; SCELT-PR = Sentence Completion Expressive Language Task–Preschool Revised; TD = typically developing; DLD = developmental language disorder.

[a]n = 7 African American children and n = 7 White children. [b]All children passed the Oral Motor subtest of the Diagnostic Evaluation of Articulation and Phonology. All children passed a binaural hearing screening at 25 dB at 1000, 2000, and 4000 kHz. [c]≤ 10th percentile.

Figure 3. Map of the brain activation areas: adult participant. Single-subject results for above-threshold activation, standard brain Montreal Neurological Institute space, $z > 2.3$ corrected to $p < .05$ using cluster correction approach.



section below. The high-resolution T1-weighted anatomical images were bias corrected, brain extracted, and normalized into Montreal Neurological Institute space using FSL (FMRIB's Software Library, https://www.fmrib.ox.ac.uk/fsl; Jenkinson et al., 2012). For the three participants acquired with a multi-echo acquisition, only the middle echo (echo time = 32 ms) was used in order to be comparable to the other participants' scans (see Table 4). Subject-level task activation maps were calculated separately for each of the volumes acquired during the sparse acquisition to account for relaxation effects.

Functional images were aligned and motion parameters were generated using FSL's *mcflirt*. Outlying volumes were determined using *fsl_motion_*outliers using the root-mean-square intensity difference metric and the default box-plot cutoff. Activation maps were generated using FSL's *fsl_glm.* The generalized linear model (GLM) included a speak > listen contrast, the six standard motion parameters as covariates, and an additional covariate for each outlying volume. The three activation maps, from the three separated acquisition volumes, were combined into a single activation map for each participant for each run using a fixed-effects analysis in

Table 4. Imaging parameters.

Sequence	Scan type	TR	TE	Field of view	Resolution	Duration (minutes:seconds)
Original protocol (n = 11)						
T1-weighted structural	3D fast field echo	8.0 ms	3.7 ms	256 × 224 × 192 mm	1 mm isotropic	6:20
SCELT-P fMRI × 2 2 sessions	Gradient echo, echo planar imaging (EPI)	2,000 ms	38 ms	240 × 240 × 128 mm	3.75 × 3.75 × 4.0 mm	14 s per trial, 30 trials[a] 3 TR + 8-s delay/trial
Updated protocol[b] (n = 3)						
T1-weighted structural	3D fast field echo	8.1 ms	3.7 ms	256 × 256 × 180 mm	1 mm isotropic	3:57
SCELT-PR fMRI × 2 2 sessions	Multi-echo EPI[c]	1,200 ms	14, 32, 50 ms	240 × 240 × 156 mm	3.0 × 3.0 × 4.0 mm	14 s per trial, 20 trials 5 TR + 8-s delay/trial

Note. TR = repetition time; TE = echo time; ms = millisecond; mm = millimeter; SCELT-P = Sentence Completion Expressive Language Task–Preschool; fMRI = functional magnetic resonance imaging; s = second; SCELT-PR = Sentence Completion Expressive Language Task–Preschool Revised.

[a]The original SCELT-P protocol was conducted using 30 trials, but due to the 30-item protocol having identical correlations as the streamlined 20-item version reported on in this study, analyses were completed using the 20 trials that overlapped between the two versions (i.e., the 20-item SCELT-PR). [b]The updated protocol made use of image acceleration to shorten overall protocol duration as well as to include more image frames in each sparse fMRI acquisition interval to allow for better characterization of the hemodynamic response. The small number of participants acquired with the new protocol did not allow for direct comparison with the earlier protocol; however, visual inspection revealed no significant artifacts in the images acquired using the accelerated protocols. [c]Multi-echo fMRI acquisition has the advantage that allows for fitting a mono-exponential decay function from which one can compute the T2* and the baseline signal (S_0) maps that separate signal changes associated with the hemodynamic response from those due to noise/motion (Kundu et al., 2017). See Maloney et al. (2023) for direct comparison of single-echo and multi-echo in an overt speech production task.

FSL's *flameo*. Then, activation maps for each run were combined into a single subject-level map using a GLM with fixed effects. Group activation maps were calculated for TD participants and participants with DLD using a GLM with a one-sample *t*-test design in FSL's *flameo*; activation maps are displayed as thresholded *z* scores.

Results

Addressing Motion During fMRI Acquisition

In principle, use of the sparse acquisition was applied to reduce motion-related artifacts. However, due to the sparse acquisition employed for the SCELT-P and SCELT-PR, motion measurement while speaking could not be directly measured since the scanner was paused during the overt response interval. That said, comparison of changes in head displacement that occurs between task conditions was possible to inform motion during fMRI acquisition. Utilizing the framewise displacement (FWD) method, employed by Power et al. (2012), we measured the change in displacement (at a 50-mm radius), from the middle volume of each scanner acquisition period during the task. The average FWD across the entire task was 4.43 mm (SD = 3.69 mm) for the DLD group (n = 5) and 2.22 mm (SD = 1.33 mm) for the TD group (n = 9). Using a two-sample independent t test, the difference between the average FWD of the DLD and TD groups is significant, $p < .03$. Significant changes were not observed in displacement between acquisition periods after speak conditions compared to listen conditions. The average FWD after a speak condition was 2.82 mm (SD = 4.19 mm), and the average after a listen condition was 3.21 mm (SD = 5.07 mm), $p < .33$, using a two-sample independent t test. As mentioned, the use of the sparse acquisition was both to mitigate the effects of head motion from a participant's speech and also to negate the interference of scanner noise on the auditory prompts.

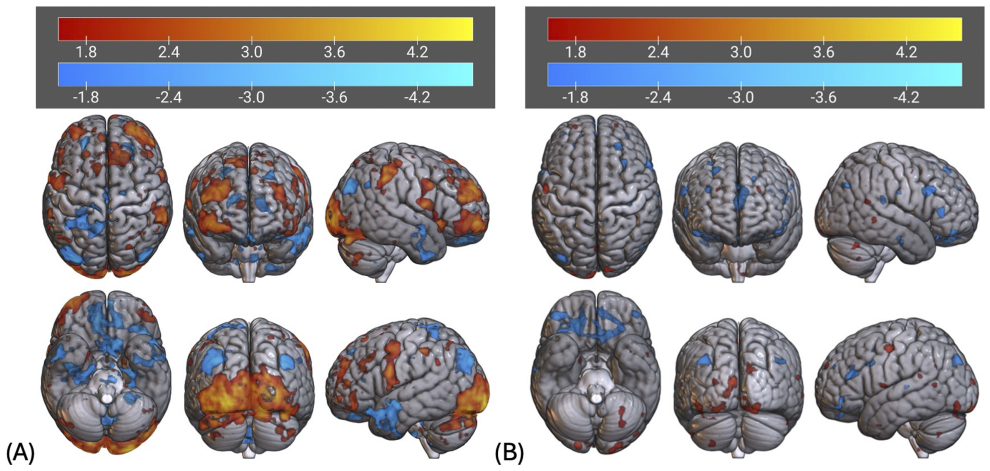
fMRI and Related Findings

Regarding structural imaging, all preschoolers (100%) successfully completed the structural imaging to visualize brain anatomy across both groups. For functional imaging, completion rates differed between groups with 89% of TD preschoolers (8/9) completing the SCELT-P (for the 20 overlapping items) during fMRI and 60% of preschoolers with DLD (3/5) successfully completing the task with limited enough motion for interpreting activation maps, free of imaging artifacts. The TD participant who was excluded from analysis successfully participated in the fMRI scan, but task performance could not be assessed due to a corrupted audio recording. One participant with DLD was excluded due to extremely high motion assessed via FWD as described above (mean FWD > 6.8 mm). Another participant with DLD was excluded due to a visible banding artifact in their activation map for the second run that suggested a technical error in imaging acquisition.

Activation

For preschoolers in the TD group who successfully completed the SCELT task (8/9), bilateral activation was observed in the inferior frontal gyrus, motor and sensory cortex, prefrontal cortex, visual cortex, and cerebellum in the experimental (listen + speak) condition versus the control (listen only) condition (see Figure 4A). The participants with DLD who successfully completed the task (3/5) exhibited more motion than their TD peers, yet the DLD group activation suggests a similar activation pattern in some regions (see Figure 4B).

Figure 4. Group composites: activation for successful completions. Map A = Group composite activation for n = 8 typically developing preschoolers; Map B = Group composite activation for n = 3 participants with developmental language disorder. For display/feasibility purposes, maps were completed with threshold at $z > 1.6$.



Discussion

The preliminary studies completed provided evidence in favor of the validity and feasibility of the SCELT-P and SCELT-PR as measures of expressive grammar for preschoolers with and without DLD. Importantly, these tasks were not designed to identify language disorder but rather were purposely designed for eliciting morphosyntactic responses beyond just the single word level within the neuroimaging environment. The preliminary findings obtained in this research note indicate evidence of strong internal consistency, excellent interrater reliability, and high test–retest reliability, confirming stability and consistency across different evaluative contexts. The significant, almost identical, correlations between the SCELT-P/R scores and established expressive grammar measures such as the CELF-P2 WS subtest, SPELT-3, and DSS metric further reinforce psychometric evidence to reflect criterion validity. Expert feedback evaluating child friendliness, computer-based administration, and the degree to which morphosyntax is targeted provided some evidence of face and content validity—necessary features for a novel measure within its early developmental stages. With the streamlined version (SCELT-PR) functioning as well as the original SCELT-P, it has demonstrated emergence as a valuable task within speech-language pathology for use when including 4- to 5-year-olds in the neuroimaging environment. It is also noteworthy to mention that the fair specificity and sensitivity, along with a cutoff score and shorter administration time, extend the SCELT-PR's utility even further, making it the recommended version for use.

Collectively, the findings observed in this study respond to earlier calls for pediatric neuroimaging tasks (cf. M. H. Lee, O'Hara, et al., 2020). Neuroimaging in the preschool years is notoriously difficult due to the tendency for young children, including those with DLD or language disorders, to exhibit significant motion artifacts (e.g., Holland et al., 2007; Hoyda et al., 2025; Sroka et al., 2015). This is especially true when these children are asked to verbalize answers while undergoing a neuroimaging scan. However, even with these challenges occurring, our findings offer preliminary evidence supporting the feasibility of including preschool-aged children in neuroimaging studies, especially as it pertains to 4- to 5-year-olds. The successful completion of imaging sessions by the majority of preschoolers in Study 2 (79%, $n = 11$ out of 14) indicates that, with careful task design and appropriate accommodations, it is possible to obtain meaningful neuroimaging data from this young age group. The caveat associated with this observation is that the external validity of our findings is limited due to the very small sample size included in the feasibility study (i.e., Study 2). However, future studies with a larger sample size of preschoolers with DLD could further this work.

The exploratory nature of Study 2 also requires a careful examination of the group maps illustrating activation (see Figures 4A and 4B) in the inferior frontal gyrus, motor and sensory cortex, prefrontal cortex, visual cortex, and cerebellum. These preliminary data suggest that similar brain activation patterns were evident across preschoolers with and without DLD. While these data are promising, neither between-group analyses nor connectivity was explored due to the small sample size, making it necessary that these findings are considered within the confines of the exploratory study design.

Limitations and Future Directions

Although this preliminary investigation offers evidence in favor of using the SCELT-PR, a few limitations are acknowledged. Validity is understood as a unitary construct supported by multiple sources of evidence, including content, construct, and criterion-related evidence (see Yusoff, 2019, for a discussion). Furthermore, with assessment validation being an iterative process that is established over time, the findings from Study 1 lay the initial groundwork for continuing to explore the validity of the SCELT-PR for use in the neuroimaging environment. In this study, we focused primarily on criterion-related evidence with considerations for face and content validity, while acknowledging that additional sources of validity evidence will further strengthen the SCELT-PR in our ongoing work.

Regarding inclusion criteria, we recognize that the cut-point of 85 on the KBIT-2 limits generalizability of nonverbal IQ scores to a subset of children with DLD. In our study, all preschoolers, including those with DLD where deficits in morphosyntax were the hallmark feature, scored at or above this cut-point. Meaning, no participants were excluded from our investigation on the basis of KBIT-2 scores. While this overly conservative criterion was intended to ensure no intellectual disability was reflected in these preschoolers' profiles, it may result in an unintended restriction of understanding the full range of cognitive–linguistic profiles tied to neuroimaging evaluations of brain structures. As such, future work should include the full range of children with DLD with nonverbal IQ scores within 2 *SD*s of the mean (cf. Bishop et al., 2017).

We also recognized as a limitation the sample sizes included in each of the two studies. In Study 1, the sample size of $N = 82$ is below the "rule-of-thumb" threshold of 10 participants per item, which limits the generalizability of the favorable psychometric evidence provided. Within Study 2, there was a small sample size of $N = 14$ for exploring feasibility within the neuroimaging environment. That said, the majority of children from both the TD (89%) and DLD (60%) groups included within Study 2 successfully engaged in the neuroimaging activities. We posit that children in this age group, especially those with DLD, would benefit from future efforts that emphasize supporting children's

health literacy and being active participants in their own care. For example, a multisensory social story geared toward children and designed to provide clear age-level expectations for required neuroimaging procedures could support children's successful participation in the neuroimaging environment (Kokotek & Washington, 2024). Although a compilation of materials was provided to families prior to completing the imaging session, a social story designed specifically for this age group and the neuroimaging activity may be of additional benefit.

Current efforts to address this limitation are underway (e.g., *Esther and the Machine*, https://www.pedllslab.com/stories or https://youtu.be/HwWzf529TiQ) in an National Institutes of Health National Institute on Deafness and Other Communication Disorders–funded (R01DC019337, Principal Investigator: Washington) randomized controlled trial (NCT05268341) across Canada and the United States with a recruitment target of 100 preschoolers, including those who are TD (n = 41) and those with hallmark morphosyntactic deficits DLD (n = 59), using the SCELT-PR. This effort will consider functional and structural connectivity to test the hypothesis that deficits in grammar learning are mediated by baseline neurobiological function (underactivation and reduced connectivity between regions in the network supporting procedural learning, e.g., inferior frontal cortex, motor cortex, basal ganglia, and cerebellum, as is articulated by the PDH). Identifying neurobiological changes underlying DLD and associated grammar learning deficits could provide insights regarding treatment targets and treatment-induced behavioral response.

An additional limitation relates to the classification accuracy analyses used in Study 1. ROC and AUC analyses were used to evaluate discriminant accuracy. However, classification accuracy estimates derived from a single sample may be influenced by sample-specific characteristics. Because the sample size did not permit cross-validation using independent subsamples, these estimates should be interpreted with some caution. Replication with larger samples will be important for confirming the stability of classification performance and further strengthening evidence supporting the SCELT-P/R's clinometrics.

Lastly, we make mention that demographic information such as maternal education, linguistic background, and race/ethnicity was not factored into the present investigation. To this point, it is important to note that the participants in this study represented a variety of backgrounds, but these characteristics were not specifically examined. Noteworthy to mention is that multilingual children were not included in this study and, for children who spoke African American English, test-specific adaptions were included to ensure interpretability and accuracy in our findings. With the strong psychometric evidence established regarding the SCELT-P/R in this preliminary investigation, additional studies that consider broader demographic features could help to ensure it is appropriate for use with preschoolers from a diverse range of backgrounds. Additionally, the psychometric merit of the SCELT-P/R may be extended in future efforts by considering item-level and test-level characteristics following this preliminary validity and feasibility effort.

Conclusions

In addition to its psychometric strengths, the SCELT-P/R holds promise as a useful task for ensuring young children are engaging in fMRI research when considering grammatical functioning beyond the single-word level. The SCELT-PR task, in particular, as the streamlined version also has the potential for providing a fruitful avenue to explore the impact of language interventions on brain function in the preschool years.

Data Availability Statement

Data included in this study are not publicly available and may only be shared using a data sharing agreement and only for parents who consented to do so. For questions, please contact the corresponding author.

Ethics Statement

Ethical approval was received from the University of Cincinnati Institutional Review Board (No. 2013-5890) in tandem with a reliance institutional review board at Cincinnati Children's Hospital Medical Center (No. 2021-0716). Research ethics board approval was also provided by the University of Toronto (No. 00043895).

Acknowledgments

This work was supported in part by the ASHFoundation (Advancing Academic-Research Careers Award, Principal Investigator [PI]: Karla N. Washington), an Imaging Research Council grant and Research Innovation Pilot Funding from the Cincinnati Children's Hospital Medical Center (Jennifer Vannest and Karla N. Washington), the U.S. National Institute on Deafness and Other Communication Disorders (Grant R01DC09337, PI: Karla N. Washington), and the Canadian Institutes of Health Research (Canada Research Chair, Grant CRC-2022-00366, PI: Karla N. Washington). We would like to thank the children and families who took part in the study, the preschools that helped with recruitment and data collection, and the practicum students and students in the Paediatric Language, Learning, and Speech (PedLLS) Outcomes Lab who supported data collection, analysis, and preparation.

References

Abbott, N., & Love, T. (2023). Bridging the divide: Brain and behavior in developmental language disorder. *Brain Sciences, 13*(11), Article 1606. https://doi.org/10.3390/brainsci13111606

Badcock, N. A., Bishop, D. V. M., Hardiman, M. J., Barry, J. G., & Watkins, K. E. (2012). Co-localisation of abnormal brain structure and function in specific language impairment. *Brain and Language, 120*(3), 310–320. https://doi.org/10.1016/j.bandl.2011.10.006

Bernstein Ratner, N., MacWhinney, B., Hong, Y., & Yang, J. S. (2026). Developmental sentence scoring for preschool language sample analysis: A psychometric update. *American Journal of Speech-Language Pathology, 35*(1), 150–169. https://doi.org/10.1044/2025_AJSLP-24-00525

Bishop, D. V. M. (1992). The underlying nature of specific language impairment. *The Journal of Child Psychology and Psychiatry, 33*(1), 3–66. https://doi.org/10.1111/j.1469-7610.1992.tb00858.x

Bishop, D. V. M. (2013). Research review: Emanuel Miller Memorial Lecture 2012—Neuroscientific studies of intervention for language impairment in children: Interpretive and methodological problems. *The Journal of Child Psychology and Psychiatry, 54*(3), 247–259. https://doi.org/10.1111/jcpp.12034

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & the CATALISE-2 Consortium. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 58*(10), 1068–1080. https://doi.org/10.1111/jcpp.12721

Brown, R. (1968). The development of wh questions in child speech. *Journal of Verbal Learning and Verbal Behavior, 7*(2), 279–290. https://doi.org/10.1016/S0022-5371(68)80002-7

Brown, R. (1973). *A first language*. Harvard University Press. https://doi.org/10.4159/9780674028920-003

Byars, A. W., Holland, S. K., Strawsburg, R. H., Bommer, W., Dunn, R. S., Schmithorst, V. J., & Plante, E. (2002). Practical aspects of conducting large-scale functional magnetic resonance imaging studies in children. *Journal of Child Neurology, 17*(12), 885–889. https://doi.org/10.1177/08830738020170122201

Chouinard, B., Boliek, C., & Jacqueline, C. (2016). How to interpret and critique neuroimaging research: A tutorial on use of functional magnetic resonance imaging in clinical populations. *American Journal of Speech-Language Pathology, 25*(3), 269–289. https://doi.org/10.1044/2016_AJSLP-15-0013

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum. https://doi.org/10.4324/9780203771587

Coltheart, M., & McArthur, G. (2012). Neuroscience, education and educational efficacy research. In M. Anderson & S. Della Sala (Eds.), *Neuroscience in education* (pp. 215–221). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199600496.003.0061

Dawson, J., Stout, C., & Eyer, J. (2003). *Structured Photographic Expressive Language Test–Third Edition*. Janelle Publications. https://www.janellepublications.com/0705.shtml

Dodd, B., Hua, Z., Crosbie, S., Holm, A., & Ozanne, A. (2006). *Diagnostic evaluation of articulation and phonology*. Harcourt Assessment. https://www.pearsonassessments.com/en-us/Store/Professional-Assessments/Speech-%26-Language/Diagnostic-Evaluation-of-Articulation-and-Phonology/p/100000295

Dunn, L. M., & Dunn, D. M. (2007). *The Peabody Picture Vocabulary Test–Fourth Edition*. Pearson. https://www.pearsonassessments.com/store/usassessments/en/Store/Professional-Assessments/Academic-Learning/Brief/Peabody-Picture-Vocabulary-Test-|-Fourth-Edition/p/100000501.html

Ebert, K. D., Ochoa-Lubinoff, C., & Holmes, M. P. (2020). Screening school-age children for developmental language disorder in primary care. *International Journal of Speech-Language Pathology, 22*(2), 152–162. https://doi.org/10.1080/17549507.2019.1632931

Ehrler, D. J., & McGhee, R. L. (2008). *Primary test of nonverbal intelligence*. Pro-Ed. https://proedinc.com/products-12590.html

Finneran, D. A., Francis, A. L., & Leonard, L. B. (2009). Sustained attention in children with specific language impairment (SLI). *Journal of Speech, Language, and Hearing Research, 52*(4), 915–929. https://doi.org/10.1044/1092-4388(2009/07-0053)

Gay, L. R. (1985). *Educational evaluation and measurement: Competencies for analysis and application*. Charles E. Merrill Publishing.

Hadley, P. (2014). Grammatical development in children with language impairment. In P. Brooks & V. Kempe (Eds.), *Encyclopedia of language development*. Sage Publications. https://doi.org/10.4135/9781483346441.n80

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8*(5), 393–402. https://doi.org/10.1038/nrn2113

Holland, S. K., Vannest, V., Mecoli, M., Jacola, L. M., Tillema, J.-M., Karunanayaka, P. R., Schmithorst, V. J., Yuan, W., Plante, E., & Byars, A. W. (2007). Functional MRI of language lateralization during development in children. *International Journal of Audiology, 46*(9), 533–551. https://doi.org/10.1080/14992020701448994

Hoyda, J., Stewart, H., Vannest, L., Washington, K. N., & More, D. R. (2025). Structural and pragmatic language skills in school-age children relate to resting state functional connectivity. *Brain Imaging and Behavior, 19*(6), 1146–1167. https://doi.org/10.1007/s11682-025-01040-7

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage, 62*(2), 782–790. https://doi.org/10.1016/j.neuroimage.2011.09.015

Justice, L. M., Ahn, W. Y., & Logan, J. A. R. (2019). Identifying children with clinical language disorder: An application of machine-learning classification. *Journal of Learning Disabilities, 52*(5), 351–365. https://doi.org/10.1177/0022219419845070

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Brief Intelligence Test–Second Edition* (KBIT-2). APA PsycTests. https://doi.org/10.1037/t27706-000

Kokotek, L., & Washington, K. N. (2024, December 5–7). *Enhancing health literacy: Using social stories to successfully include children in the neuroimaging environment* [Technical research session - virtual]. American Speech-Language-Hearing Association Convention, Seattle, WA, United States of America.

Krishnan, S., Cler, G. J., Smith, H. J., Willis, H. E., Asaridou, S. S., Healy, M. P., Papp, D., & Watkins, K. E. (2022). Quantitative MRI reveals differences in striatal myelin in children with DLD. *eLife, 11,* Article e74242. https://doi.org/10.7554/eLife.74242

Kundu, P., Voon, V., Balchandani, P., Lombardo, M. V., Poser, B. A., & Bandettini, P. A. (2017). Multi-echo fMRI: A review of applications in fMRI denoising and analysis of BOLD signals. *NeuroImage, 154,* 59–80. https://doi.org/10.1016/j.neuroimage.2017.03.033

Lee, J. C., Dick, A. S., & Tomblin, J. B. (2020). Altered brain structures in the dorsal and ventral language pathways in individuals with and without developmental language disorder (DLD). *Brain Imaging and Behavior, 14*(6), 2569–2586. https://doi.org/10.1007/s11682-019-00209-1

Lee, L. (1974). *Developmental sentence analysis: A grammatical assessment procedure for speech and language clinicians*. Northwestern University Press.

Lee, L. L., & Canter, S. M. (1971). Developmental sentence scoring: A clinical procedure for estimating syntactic development in

children's spontaneous speech. *Journal of Speech and Hearing Disorders, 36*(3), 315–340. https://doi.org/10.1044/jshd.3603.315

Lee, M. H., O'Hara, N. B., Behen, M. E., & Jeong, J. W. (2020). Altered efficiency of white matter connections for language function in children with language disorder. *Brain and Language, 203,* Article 104743. https://doi.org/10.1016/j.bandl.2020.104743

Leon, M., Washington, K. N., Fritz, K. A., Leon, M., Basinger, M., & Crowe, K. (2021). Intelligibility in Context Scale: Sensitivity and specificity in the Jamaican context. *Clinical Linguistics & Phonetics, 35*(2), 154–171. https://doi.org/10.1080/02699206.2020.1766574

Leonard, L. B. (2013). Specific language impairment across languages. *Child Development Perspectives, 8*(1), 1–5. https://doi.org/10.1111/cdep.12053

Leonard, L. B., Camarata, S. M., Pawlowska, M., Brown, B., & Camarata, M. N. (2006). Tense and agreement morphemes in the speech of children with specific language impairment during intervention: Phase 2. *Journal of Speech, Language, and Hearing Research, 49*(4), 749–770. https://doi.org/10.1044/1092-4388(2006/054)

Leonard, L. B., & Deevy, P. (2020). Retrieval practice and word learning in children with specific language impairment and their typically developing peers. *Journal of Speech, Language, and Hearing Research, 63*(10), 3252–3262. https://doi.org/10.1044/2020_JSLHR-20-00006

Maloney, T. C., Dietz, A., Vannest, J., Wilkinson, K., Szaflarski, J. P., Stall, C., & Mamlekar, C. R. (2023). Functional magnetic resonance imaging activation during unscripted discourse in people with poststroke aphasia. *Journal of Speech, Language, and Hearing Research, 66*(12), 4838–4848. https://doi.org/10.1044/2023_JSLHR-23-00124

Marchman, V. A., Wulfeck, B., & Weismer, S. E. (1999). Morphological productivity in children with normal language and SLI. *Journal of Speech, Language, and Hearing Research, 42*(1), 206–219. https://doi.org/10.1044/jslhr.4201.206

Mayes, A. K., Reilly, S., & Morgan, A. T. (2015). Neural correlates of childhood language disorder: A systematic review. *Developmental Medicine & Child Neurology, 57*(8), 706–717. https://doi.org/10.1111/dmcn.12714

McGregor, K. K. (2020). How we fail children with developmental language disorder. *Language, Speech, and Hearing Services in Schools, 51*(4), 981–992. https://doi.org/10.1044/2020_LSHSS-20-00003

McGregor, K. K., Goffman, L., Plante, E., & Werfel, K. (2025). Developmental language disorder as a multidimensional neurodevelopmental spectrum: Implications for diagnosis. *Journal of Speech, Language, and Hearing Research, 68*(12), 6069–6088. https://doi.org/10.1044/2025_JSLHR-25-00089

Messick, S. (1993). Foundations of validity: Meaning and consequences in psychological assessment. *ETS Research Report Series*, 1993(2), i–18. https://doi.org/10.1002/j.2333-8504.1993.tb01562.x

Miller, J. F. (1996). Progress in assessing, describing, and defining child language disorder. In K. N. Cole, P. S. Dale, & D. J. Thal (Eds.), *Assessment of communication and language* (Vol. 6, pp. 309–324). Brookes.

Ning, R., Bazzocchi, N., & Washington, K. (2025, June 24–27). Neuroimaging as a feasible tool for studying late talking in toddlers [Paper presentation]. International Clinical Phonetics and Linguistics Association (ICPLA), Patras, Greece.

Nunnally, J. C., & Bernstein, I. H. (1994). The assessment of reliability. *Psychometric Theory, 3,* 248–292.

Pace, A., Curran, M., Van Horne, A. O., de Villiers, J., Iglesias, A., Golinkoff, R. M., Wilson, M. S., & Hirsh-Pasek, K. (2022). Classification accuracy of the Quick Interactive Language Screener for preschool children with and without developmental language disorder. *Journal of Communication Disorders, 100,* Article 106276. https://doi.org/10.1016/j.jcomdis.2022.106276

Paul, R., & Alforde, S. (1993). Grammatical morpheme acquisition in 4-year-olds with normal, impaired, and late-developing language. *Journal of Speech and Hearing Research, 36*(6), 1271–1275. https://doi.org/10.1044/jshr.3606.1271

Paul, R., & Norbury, C. F. (2012). *Language disorders from infancy through adolescence: Listening, speaking, reading, writing, and communicating* (4th ed.). Elsevier Mosby.

Perona, K., Plante, E., & Vance, R. (2005). Diagnostic accuracy of the Structured Photographic Expressive Language Test: Third Edition (SPELT-3). *Language, Speech, and Hearing Services in Schools, 36*(2), 103–115. https://doi.org/10.1044/0161-1461(2005/010)

Pigdon, L., Willmott, C., Reilly, S., Conti-Ramsden, G., Liégeois, F., Connelly, A., & Morgan, A. T. (2020). The neural basis of nonword repetition in children with developmental speech or language disorder: An fMRI study. *Neuropsychologia, 138,* Article 107312. https://doi.org/10.1016/j.neuropsychologia.2019.107312

Plante, E., Patterson, D., Sandoval, M., Vance, C. J., & Asbjornsen, A. E. (2017). An fMRI study of implicit language learning in developmental language impairment. *NeuroImage: Clinical, 14,* 277–285. https://doi.org/10.1016/j.nicl.2017.01.027

Plante, E., & Vance, R. (1994). Selection of preschool language tests: A data-based approach. *Language, Speech, and Hearing Services in Schools, 25*(1), 15–24. https://doi.org/10.1044/0161-1461.2501.15

Portney, L. G., & Watkins, M. P. (2009). *Foundations of clinical research: Applications to practice* (3rd ed.). Pearson Education, Inc.

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage, 59*(3), 2142–2154. https://doi.org/10.1016/j.neuroimage.2011.10.018

Schmithorst, V. J., & Holland, S. K. (2004). Event-related fMRI technique for auditory processing with hemodynamics unrelated to acoustic gradient noise. *Magnetic Resonance in Medicine, 51*(2), 399–402. https://doi.org/10.1002/mrm.10706

Sealey, L. R., & Gilmore, S. E. (2008). Effects of sampling context on the finite verb production of children with and without delayed language development. *Journal of Communication Disorders, 41*(3), 223–258. https://doi.org/10.1016/j.jcomdis.2007.10.002

Shipley, K. G., Stone, T. A., & Sue, M. B. (1983). *Test for Examining Expressive Morphology*. Communication Skill Builders.

Sroka, M., Vannest, J., Maloney, T., Horowitz-Kraus, T., Byars, A., Holland, S., & CMIND Authorship Consortium. (2015). Relationship between receptive vocabulary and the neural substrates for story processing in preschoolers. *Brain Imaging Behavior, 9*(1), 43–55. https://doi.org/10.1007/s11682-014-9342-8

Streiner, D. L., & Norman, G. R. (2008). *Health measurement scales: A practical guide to their development and use* (4th ed.). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199231881.001.0001

Streiner, D. L., Norman, G. R., & Cairney, J. (2014). *Health Measurement Scales: A practical guide to their development and use* (5th ed.). Oxford University Press. https://doi.org/10.1093/med/9780199685219.001.0001

Uddén, J., Hultén, A., Schoffelen, J.-M., Lam, N., Harbusch, K., van den Bosch, A., Kempen, G., Petersson, K. M., & Hagoort, P. (2022). Supramodal sentence processing in the human

brain: fMRI evidence for the influence of syntactic complexity in more than 200 participants. *Neurobiology of Language, 3*(4), 575–598. https://doi.org/10.1162/nol_a_00080

Ullman, M. T., Clark, G. M., Pullman, M. Y., Lovelett, J. T., Pierpont, E. I., Jiang, X., & Turkeltaub, P. E. (2024). The neuroanatomy of developmental language disorder: A systematic review and meta-analysis. *Nature Human Behaviour, 8*(5), 962–975. https://doi.org/10.1038/s41562-024-01843-6

Ullman, M. T., Earle, F. S., Walenski, M., & Janacsek, K. (2020). The neurocognition of developmental disorders of language. *Annual Review of Psychology, 71*(1), 389–417. https://doi.org/10.1146/annurev-psych-122216-011555

Valian, V. (1992). Categories of first syntax: BE, BE + ING, and nothingness. In J. M. Meisel (Ed.), *The acquisition of verb placement*. Kluwer Academic Publishers. https://doi.org/10.1007/978-94-011-2803-2_13

Valian, V. (2006). Young children's understanding of present and past tense. *Language Learning and Development, 2*(4), 251–276. https://doi.org/10.1207/s15473341lld0204_2

Vannest, J., Rajagopal, A., Cicchino, N. D., Franks-Henry, J., Simpson, S. M., Lee, G. R., Altaye, M., Sroka, C., & Holland, S. K. (2014). Factors determining success of awake and asleep magnetic resonance imaging scans in nonsedated children. *Neuropediatrics, 45*(6), 370–377. https://doi.org/10.1055/s-0034-1387816

Verly, M., Gerrits, R., Sleurs, C., Lagae, L., Sunaert, S., Zink, I., & Rommel, N. (2019). The mis-wired language network in children with developmental language disorder: Insights from DTI tractography. *Brain Imaging and Behavior, 13*(4), 973–984. https://doi.org/10.1007/s11682-018-9903-3

Vydrova, R., Komarek, V., Sanda, J., Sterbova, K., Jahodova, A., Maulisova, A., Zackova, J., Reissigova, J., Krsek, P., & Kyncl, M. (2015). Structural alterations of the language connectome in children with specific language impairment. *Brain and Language, 151,* 35–41. https://doi.org/10.1016/j.bandl.2015.10.003

Washington, K. N., Murrison, M., McDonald, M., Waldo, C., Westhoven, M., Henchy, C., Maloney, T., & Vannest, J. (2015, November). *Feasibility of a Valid and Reliable Computer-Based Expressive Language Task for 4-to-5 Year Old Children* [Oral presentation]. American Speech-Language-Hearing Association Convention, Denver, Colorado, United States.

Washington, K. N., Murrison, M., Vannest, J., & Holland, S. (2013). *Sentence Completion Expressive Language Task–Preschool (SCELT-P): An fMRI task for preschoolers*. Moodoo Productions.

Washington, K. N., Murrison, M., Vannest, J., & Holland, S. (2014, November). *Validity and reliability of a computer-based expressive language task for 4- to 5-year olds* [Technical research session]. In Proceedings of the American Speech-Language-Hearing Association Convention, Orlando, Florida, United States.

Washington, K. N., Redle, R., Murrison, M., & Vannest, J. (2018). *Sentence Completion Expressive Language Task–Preschool Revised (SCELT-PR): An fMRI task for preschoolers*. Moodoo Productions.

Washington, K. N., Warr-Leeper, G., & Thomas-Stonell, N. (2011). Exploring the outcomes of a novel computer-assisted treatment program targeting expressive-grammar deficits in preschoolers with SLI. *Journal of Communication Disorders, 44*(3), 315–330. https://doi.org/10.1016/j.jcomdis.2011.01.002

Washington, K. N., & Warr-Leepr, G. (2013). Visual support in intervention for preschoolers with specific language impairment. *Topics in Language Disorders, 33*(4), 347–365. https://doi.org/10.1097/01.TLD.0000437941.08860.2f

Westby, C. (2007). Application of the ICF in children with language impairments. *Seminars in Speech and Language, 28*(4), 265–272. https://doi.org/10.1055/s-2007-986523

Wiig, E. H., Secord, W. A., & Semel, E. (2004). *Clinical Evaluation of Language Fundamentals Preschool–Second Edition*. Harcourt Assessment.

Williams, K. T. (2007). *EVT-2: Expressive Vocabulary Test–Second Edition*. Pearson. https://doi.org/10.1037/t15094-000

Yarian, M., Washington, K. N., Spencer, C. E., Vannest, J., & Crowe, K. (2021). Exploring predictors of expressive grammar across different assessment tasks in preschoolers with or without DLD. *Communication Disorders Quarterly, 42*(2), 111–121. https://doi.org/10.1177/1525740119868238

Yusoff, M. S. B. (2019). ABC of content validation and content validity index calculation. *Education in Medicine Journal, 11*(2), 49–54. https://doi.org/10.21315/eimj2019.11.2.6