JSLHR

Research Article

Structural Relationship Between Cognitive Processing and Syntactic Sentence Comprehension in Children With and Without Developmental Language Disorder

James W. Montgomery,[a] Julia L. Evans,[b] Jamison D. Fargo,[c] Sarah Schwartz,[c] and Ronald B. Gillam[d]

Purpose: We assessed the potential direct and indirect (mediated) influences of 4 cognitive mechanisms we believe are theoretically relevant to canonical and noncanonical sentence comprehension of school-age children with and without developmental language disorder (DLD).
Method: One hundred seventeen children with DLD and 117 propensity-matched typically developing (TD) children participated. Comprehension was indexed by children identifying the agent in implausible sentences. Children completed cognitive tasks indexing the latent predictors of fluid reasoning (FLD-R), controlled attention (CATT), complex working memory (cWM), and long-term memory language knowledge (LTM-LK).
Results: Structural equation modeling revealed that the best model fit was an indirect model in which cWM mediated the relationship among FLD-R, CATT, LTM-LK, and sentence comprehension. For TD children, comprehension of both sentence types was indirectly influenced by FLD-R (pattern recognition) and LTM-LK (linguistic chunking). For children with DLD, canonical sentence comprehension was indirectly influenced by LTM-LK and CATT, and noncanonical comprehension was indirectly influenced just by CATT.
Conclusions: cWM mediates sentence comprehension in children with DLD and TD children. For TD children, comprehension occurs automatically through pattern recognition and linguistic chunking. For children with DLD, comprehension is cognitively effortful. Whereas canonical comprehension occurs through chunking, noncanonical comprehension develops on a word-by-word basis.
Supplemental Material: https://doi.org/10.23641/asha.7178939

We have been engaged in a two-pronged research program, one centering on better understanding the syntactic comprehension of children with developmental language disorder (DLD)[1] and the other centering on the role that cognitive processing plays in children's syntactic comprehension. Investigation of the syntactic comprehension of these children is driven by the need to build comprehension frameworks from the ground up—frameworks that include greater specificity about the importance of individual cues and the collection of cues (e.g., syntactic, semantic, pragmatic). Our motivation for studying the relationship between cognitive processing and sentence comprehension derives from the absence of theoretically, empirically, and developmentally grounded models, which has led to a shallow understanding of the nature of sentence comprehension in DLD.

Our research program follows a psycholinguistic tradition in which sentence comprehension reflects the intersection of language knowledge/processing and general cognitive processing abilities such as fluid (analytic) reasoning, controlled attention (CATT), and memory. Spoken sentence comprehension represents a unique problem-solving activity as listeners must make immediate sense of a rapidly

[a] Division of Communication Sciences and Disorders, Ohio University, Athens
[b] School of Behavioral and Brain Sciences, University of Texas–Dallas, Richardson
[c] Department of Psychology, Utah State University, Logan
[d] Department of Communicative Disorders and Deaf Education, Utah State University, Logan

Correspondence to James W. Montgomery: montgoj1@ohio.edu

Editor-in-Chief: Sean Redmond
Editor: Lisa Archibald

Received November 11, 2017
Revision received April 2, 2018
Accepted June 29, 2018
https://doi.org/10.1044/2018_JSLHR-L-17-0421

---

[1] In this article, we use the term *developmental language disorder,* similar to the terms *primary language impairment* and *language learning impairment,* as being synonymous with the broad clinical definition of *specific language impairment* that encompasses both language and cognitive processing deficits.

Disclosure: *Ron Gillam receives royalties from the sale of the Test of Narrative Language, which was administered to participants. No other competing interests existed at the time of publication.*

disappearing signal. Listeners must establish and maintain intermediate linguistic structures in an active mental state while creating new structures and combining the sequence of structures into a single, coherent representation of the entire sentence. Fluid reasoning (FLD-R) is the general ability to recognize novel patterns and sequences of relationships and to reason about the relationships in the service of making a decision (Haavisto & Lehto, 2005). It has been suggested that FLD-R and syntactic processing are, to some extent, similar because both involve recognizing and interpreting patterns in the input (Andrews, Ogden, & Halford, 2017). CATT is a complex concept that encompasses several different abilities, with two of them being sustained attention and attention switching. These attention abilities may be relevant to comprehension as listeners attend to the words of a speaker and switch their attention between the input and making meaning of the words. Memory is a complex construct often divided broadly into two subsystems: working memory (WM) and long-term memory (LTM). WM enables individuals to engage in simultaneous information processing and storage. We assume that WM should play an important role in comprehension because, at its core, understanding a sentence involves the ability to store linguistic representations that have already been generated while new ones are created from downstream material. LTM reflects an individual's permanent knowledge base. Language (crystallized) knowledge resides in LTM and includes phonological, morphological, lexical, and syntactic representations and should guide listeners in making sense of the input (Kintsch, Patel, & Ericsson, 1999; Kutas & Federmeier, 2000).

Sentence Comprehension in School-Age Children With DLD: A Snapshot

Sentence comprehension deficits are a hallmark feature of school-age children with DLD. Relative to typically developing (TD) children, children with DLD demonstrate poor comprehension of noncanonical sentences such as reversible *be* passives (*The lion* $_{[i]}$ *was bitten* $_{[\mathbf{ti}]}$ *by the monkey*; Montgomery & Evans, 2009; Norbury, Bishop, & Briscoe, 2002; van der Lely, 1998; van der Lely & Stollwerck, 1997). They also have marked difficulty in understanding object relatives (*The lion* $_{[i]}$ *that the monkey bit* $_{[\mathbf{ti}]}$ *was brown*; Friedmann & Novogrodsky, 2004, 2007).

From a linguistic perspective, passives and object relatives are especially hard for children with DLD because they violate the noun–verb–noun (NVN)/subject–verb–object (SVO) word order of English in which noun phrase 1 (NP1) is the agent and noun phrase 2 (NP2) is the patient (*The monkey bit the lion*). Even though the surface form of passives is NVN and in object relatives it is NNV, children must come to realize that in both structures (a) NP1 appears in subject position but functions as the patient and (b) NP2 appears in object position but functions as the agent. In both cases, there is movement of the logical object–noun phrase (NP; *the lion*) to the subject position. The canonical relationship of the object to its verb (*bitten, bit*) is maintained by the moved element, leaving a trace ([t*i*]) in its original object position. The trace shares a coreferential relation with the moved element ([*i*]). Reactivation of the moved element is presumably required in both structures to establish a filler-gap dependency, with NP1 (filler, marked as [*i*]) being reactivated after encountering the verb (trace/gap [t*i*]). Following reactivation, NP2 is assigned the agent role and NP1 is assigned the patient role, allowing the listener to determine who did what to whom.

Canonical structures, though, can also be difficult for children with DLD. Compared with TD peers, children with DLD reliably show poorer comprehension of SVO-like structures when they become lengthy and contain lexical detail (*The yellow dog washes the white pig; The short fat clown is hugging the blonde-haired girl*) that is important to distinguishing the agent and patient of the sentence (Leonard, Deevy, Fey, & Bredin-Oja, 2013; Montgomery, Evans, & Gillam, 2009; Robertson & Joanisse, 2010). Such findings implicate nonsyntactic factors in the sentence comprehension difficulties of these children.

Influence of Cognitive Processing on Sentence Comprehension

Children with DLD exhibit limitations in a variety of cognitive abilities such as phonological short-term memory (Archibald & Gathercole, 2007), verbal complex WM (cWM; Archibald & Gathercole, 2007; Ellis Weismer, Evans, & Hesketh, 1999), and CATT (Lum, Conti-Ramsden, & Lindell, 2007; Marton, Campanelli, Eichorn, Scheuer, & Yoon, 2014). Our understanding of the relationship between cognitive processing and sentence comprehension in DLD is far from clear. Below, we focus on a review of the influence of (a) verbal cWM because it involves concurrent verbal processing and storage and (b) CATT. We then turn to the potential roles of language LTM and FLD-R because of their theoretical relevance to comprehension. Collectively, these four cognitive mechanisms are most relevant to the current study.

cWM and sentence comprehension. Studies examining the association between cWM and sentence comprehension in children with DLD have assumed these children have trouble building ongoing structure and making proper semantic role assignments while remembering the products of earlier processing (Just & Carpenter, 1992; Just, Carpenter, & Keller, 1996). Emerging evidence suggests that, for children with DLD, verbal cWM relates to the comprehension of verbal *be* passives (Montgomery & Evans, 2009) and lengthy SVO structures (Montgomery, 2000; Montgomery et al., 2009). Other studies with TD children have shown a relation between verbal cWM and the comprehension of passives and object relatives (Ahmad Rusli & Montgomery, 2017; Montgomery, Magimairaj, & O'Malley, 2008; Weighall & Altmann, 2011).

CATT and sentence comprehension. Sustained attention has been examined as a potential influence on DLD sentence comprehension, with the assumption being that the ability to maintain attention over the course of a sentence should relate to comprehension. There is evidence showing that sustained auditory attention relates to the real-time sentence processing and offline comprehension of children with DLD, but not same-age peers (Montgomery, 2008; Montgomery et al., 2009). Such findings suggest that the comprehension of even simple grammar by children with DLD is not yet automatic. There is also evidence implying that attention switching is important to TD children's sentence comprehension (Finney, Montgomery, Gillam, & Evans, 2014). Children listened to object relative sentences (*The goat* $_{[i]}$ *that the pig had bumped* $_{[ti]}$ *near the bush was smiling*) and selected the agent of the sentence from two images (agent, patient) presented after the embedded verb/syntactic gap *[*$_{ti}$*]*. It was reasoned that the children should redirect their attention to WM storage to reactivate NP1 [i] after processing the embedded verb to create long-distance syntactic dependency.

Influences of language LTM and FLD-R on DLD sentence comprehension. Studies explicitly designed to investigate the influence of language LTM on DLD sentence comprehension are nonexistent. Likewise, there is an absence of studies examining the relation of FLD-R and sentence comprehension.

Language LTM and sentence comprehension. Within the TD literature, results of two studies (Alonzo, Yeomans-Maldonado, Murphy, & Bevens, 2016; Boyle, Lindell, & Kidd, 2013), not surprisingly, imply that language knowledge (indexed by sentence repetition) predicted children's sentence comprehension. The language knowledge and processing schemes stored in LTM used to repeat sentences are the same that support the construction of a sentence representation from a string of input words.

FLD-R and sentence comprehension. It has been suggested that FLD-R and syntactic processing are similar to some extent as both involve recognizing and interpreting underlying structures or patterns in the input (Andrews et al., 2017). The implausible sentences in the current study represented a novel problem-solving situation because they forced the children to rely on their syntactic knowledge to determine which noun functioned as the agent and which noun functioned as the patient. Given the novelty of our sentences, we reasoned that children's general pattern recognition/analytic abilities may play a role in comprehension.

We are unaware of any studies examining the relation of FLD-R and comprehension in children with DLD. However, two studies with adults are relevant here. Andrews et al. (2017) showed that adults with stronger FLD-R were better at determining the agent–action relationships in syntactically complex but implausible sentences than those with weaker FLD-R abilities. Moreover, FLD-R accounted for unique variance in comprehension even after controlling for cWM. The authors argued that FLD-R allowed participants to resolve the conflicts between syntactic structure and semantic plausibility. Engelhardt, Nigg, and Ferreira (2017) also reported a significant relation between FLD-R and syntactic comprehension in adults. Closer to the aim of this study, FLD-R has been shown to relate to the reading comprehension of school-age TD children (García-Madruga, Vila, Gómez-Veiga, Duque, & Elosúa, 2014; Motallebzadeh & Yazdi, 2016).

Limitations to Our Understanding of DLD Sentence Comprehension

We see two major shortcomings in the DLD literature that have hampered our understanding the nature of the sentence comprehension abilities of children with DLD. Although adult language researchers (e.g., Bates & MacWhinney, 1989; Kim & Sikos, 2011; Kutas & Federmeier, 2000) and developmental researchers (e.g., Bates & MacWhinney, 1987, 1989) have studied the influence of various syntactic and semantic–pragmatic cues on comprehension, DLD researchers have not. Consequently, linguistic frameworks detailing the influence of various cues on the comprehension of these children are lacking. Understanding the cues used by these children will provide us important linguistic insights into how these children come to comprehend what they hear.

Word order and animacy in English are powerful cues guiding comprehension. Given that agency and animacy tend to co-occur (Kuperberg, Kreher, Sitnikova, Caplan, & Holcomb, 2007), the semantic role of agent is typically assigned to the animate noun that has a predictable relationship to the action and the patient (*The dog chased the cat; The cat was chased by the dog*). Within the DLD literature, studies by Dick, Wulfeck, Krupa-Kwiatkowski, and Bates (2004) and Evans (Evans, 2002; Evans & MacWhinney, 1999) are notable as being the few studies to try to disentangle the influences of different cues (i.e., word order and animacy) on DLD sentence comprehension. Following this line of research, Montgomery, Gillam, Evans, and Sergeev (2017) offer some clarity on the use of a single cue—word order—to guide comprehension in school-age (7- to 11-year-old) children with DLD. Children were asked to comprehend highly implausible sentences, which forced them to rely solely on word order to comprehend the sentences. Compared with same-age peers, children with DLD exhibited poorer comprehension of canonical and noncanonical sentences, especially noncanonical, implying that children with DLD have significant trouble acquiring/using word order knowledge, with noncanonical word order posing special problems.

A second limitation is that previous studies exploring the relationship between cognitive processing and sentence comprehension have (a) employed small samples; (b) exercised different experimental controls in the construction of sentence stimuli; (c) investigated a limited range of cognitive mechanisms, usually a single mechanism within a given study; and (d) not been designed to examine direct and/or indirect influences of various cognitive mechanisms on comprehension. This study was designed to overcome these shortcomings.

Theoretical Framework and Overview of This Study

The current study was designed to determine the structural relationship between cognitive processing and sentence comprehension in young school-age children with and without DLD. We studied 7- to 11-year-old children because this is the age range during which this relationship has been examined with some intensity, thus offering us some guidance about which cognitive mechanisms might be the most theoretically relevant to study. The mechanisms of interest included the following: FLD-R, reflecting children's novel pattern recognition and problem-solving/analytic reasoning abilities; CATT, including sustained attention and attention switching; cWM, representing children's ability to engage in concurrent verbal processing and storage; and language knowledge in long-term memory (LTM-LK) that includes children's crystallized language knowledge. A simple relationship would be implied if each mechanism has its own direct influence on comprehension. A more complex relationship would be reflected by one or more of the mechanisms playing a mediating role through which the other mechanisms indirectly influence comprehension. We predicted a complex relationship (see below).

We envisioned these mechanisms to be influential in comprehension in the following way. First, FLD-R should be important because our implausible sentences represent a novel problem for the children to solve (i.e., understand who did what to whom), leading us to reason that children may use their general pattern recognition/analytic reasoning abilities to help solve the problem. Second, LTM-LK should be important as comprehension relies on the ability to recognize word order patterns and assign meaning to those patterns. As children listen to a sentence they activate from LTM lexical items and the associated syntactic categories (noun, verb, etc.) of those items as well as multiword representations corresponding to the distributional characteristics of the input (see below for greater detail on this assumption). Third, sustained attention and attention switching may be important as they allow children to attend to incoming lexical items and toggle their attention between processing the incoming words and storing the products of processing. Fourth, cWM should play a pivotal role as it is the principal function of cWM (concurrent processing and storage) that should enable children to hold in an active state structures generated from earlier processing while building new structures from downstream material (e.g., Just & Carpenter, 1992; Just et al., 1996). It is precisely this dual function that leads us to believe that cWM should be the mediating mechanism through which FLD-R, CATT, and language LTM indirectly influence comprehension.

Our assumption that cWM should play the mediating role also leads us to consider the embedded processes model of cWM proposed by Cowan and associates (Cowan, 1999; Cowan, Rouder, Blume, & Saults, 2012; Cowan, Saults, & Blume, 2014) as being especially relevant to comprehension. Its relevance lies in its parsimonious view of storage and LTM—an integrated view of attention and cWM storage—and the function of chunking in LTM, which allows listeners to consolidate a string of incoming words into fewer, coherent units, thereby maximizing storage. This model emphasizes that those items that are important to accomplishing some goal (e.g., sentence comprehension) are activated from LTM and that these activated items occupy central storage and peripheral storage. Central storage (focus of attention) is limited to about one item (e.g., phrase, clause), and peripheral storage contains the remaining activated items that lie just outside the focus of attention. The total capacity of cWM is the sum of central and peripheral storage, which is limited to about three to five items or chunks. Importantly, items in cWM may be of variable size, depending on whether they have been chunked or not. Chunking is a key functional feature of the model because it enables individuals to consolidate many initially encountered items into fewer, larger, but integrated units. Attention plays a "zooming" role in that it initially zooms out to capture several items during the encoding of the input (Cowan et al., 2005) and then zooms in to maintain just one item in central storage at any given moment. Central storage and focal attention are considered one and the same (Cowan, 1999). Cowan, Fristoe, Elliott, Brunner, and Saults (2006), however, also acknowledge that the broader view of attention (i.e., CATT) is important to cWM, for example, with sustained/selective attention used to maintain items in memory (Unsworth & Engle, 2007). Apart from the Cowan model, attention switching appears to be another important mechanism influencing cWM (Barrouillet, Gavens, Vergauwe, Gaillard, & Camos, 2009; Unsworth & Robison, 2015, 2016, 2017), allowing individuals to rapidly alternate focal attention between storage (to maintain items in storage) and the processing component of the task.

From Cowan's perspective, we would make the following arguments about the intertwined roles of cWM, LTM, and CATT in relation to sentence comprehension. First, lexical items in a spoken sentence are activated from LTM. Language content (phonology, morphology, semantics, syntax) represents the listener's long-term language knowledge. Lexical–semantic knowledge is considered part of declarative LTM, whereas syntactic knowledge is part of procedural LTM (Ullman, 2001, 2004). Syntactic knowledge allows the creation of intermediate structures–meanings (phrases, clauses) and a final sentence representation by combining the intermediate structures. Second, activated words are incrementally and automatically chunked by the language system into units such as phrases and clauses and then combined into a sentence representation (Gilchrist, Cowan, & Naveh-Benjamin, 2009). Third, given the severe storage limits of cWM (Cowan et al., 2014; McElree, 2006, Oberauer & Hein, 2012), central storage should contain the most recent chunk (e.g., phrase, partial or complete clause) whereas peripheral storage holds those chunks that have been created from earlier processing. The language system then combines these chunks into a sentence representation, allowing comprehension to take place.

With respect to CATT, we might contend that sustaining attention over the course of a sentence should allow listeners to attend to the incoming words of a sentence, thereby promoting the chunking of input into relevant linguistic units. Note here that we do not assume sustained attention plays an item maintenance function (given the severe time constraints inherent in comprehension), but instead a permission function allowing listeners to attend to incoming lexical input. Attention switching may also be important if we assume that comprehension requires listeners to rapidly toggle focal attention between cWM, where the linguistic chunks that have already been created are stored, and the language system in order to generate new chunks and to combine them with previous chunks into a full sentence representation (Finney et al., 2014).

The Challenges of Comprehension and Two Possible Solutions: Potential Relevance to the Current Study

To understand a spoken sentence, a listener incrementally builds and combines intermediate structures and meaning into an integrated sentence representation. But listeners are faced with two fundamental challenges. A sentence representation must be built immediately from a fleeting signal, and cWM is severely constrained by the number of chunks it can hold in support of comprehension.

Two recent and complementary conceptualizations of adult sentence processing/comprehension have been forwarded to address these challenges. Both imply a strong link between cWM and LTM, and both appear to have relevance to our thinking about the sentence comprehension of children. The first, rooted in a connectionist view of language (Christiansen & MacDonald, 1999) and usage-based view (Abbot-Smith & Tomasello, 2006; Lieven, Salomo, & Tomasello, 2009), comes from the work of Christiansen and colleagues (Arnon & Christiansen, 2017; Chater, McCauley, & Christiansen, 2016; Christiansen & Arnon, 2017; Christiansen & Chater, 2016; McCauley & Christiansen, 2013, 2017). These authors propose a “chunk-and-pass” model based on the authors’ view that language learning is a skill rooted in learning “how” to process language. Syntactic structure emerges as a natural byproduct of repeated experiences processing language, which enables children to create memory traces of structures to be used for future processing. A key processing principle of the model is rapid input chunking and passing chunks created at lower levels (e.g., phonological, lexical) to higher levels (e.g., multiword units, syntactic). Immediately chunking input into multiword units allows the creation of intermediate and more abstract structures (NPs, verb phrases, clauses), which, importantly, reduces the online memory load on the listener because the chunks can be maintained over a broad time window as new information is chunked. Chunking occurs iteratively over the course of the input until all necessary structures are developed, at which point they are combined into a single, coherent structure. As structures are built, the listener uses available semantic–pragmatic cues to assign meaning to the components of the structures. Chunking is facilitated as the language system accrues more multiword representations that can then be automatically activated and used during future input processing. Adults are sensitive to multiword units during spoken sentence comprehension (Arnon & Snider, 2010; Reali & Christiansen, 2007) and reading comprehension (Tremblay, Derwing, Libben, & Westbury, 2011). Individual differences in chunking ability have been shown to predict variation in adults’ online processing of complex sentences (McCauley & Christiansen, 2015).

Evidence in the developmental literature is emerging to suggest that multiword chunks are important building blocks for syntactic development and that such chunks are used for comprehension and production (Arnon & Clark, 2011; Arnon, McCauley, & Christiansen, 2017; Bannard & Lieven, 2012; Bannard & Matthews, 2008; Cornish, Dale, Kirby, & Christiansen, 2017; McCauley & Christiansen, 2013, 2014; McCauley, Isbilen, & Christiansen, 2017). As already noted, repeated language processing experiences enable children to acquire and store multiword representations, which provides them crucial information about the distributional and structural relationships across words and the opportunity to reuse the component parts of the units (Ns, Vs, NV) to acquire new multiword units (Cornish et al., 2017; Theakston & Lieven, 2017).

Applied to the current study, children would be expected to first learn the basic canonical NVN structure (*The boy kissed the girl*) at a very early age (Bencini & Valian, 2008; Yuan & Fisher, 2009; Yuan, Fisher, & Snedeker, 2012). From this simple frequently occurring structure, children reuse the familiar N, V, N components and/or the multiword NVN sequence to acquire subject relative forms (*The boy that kissed the girl…*). By contrast, the learning of the noncanonical passive structure (*The girl was kissed by the boy*) not only requires children to recognize a familiar NVN chunk but also, most crucially, that two intervening bits of information occur between N1 and N2, the past tense *–ed* morpheme on the verb and a *by*-phrase, both of which must be learned as reliable cues marking the input as a *be* passive. In the case of object relative sentences (*The girl that the boy kissed was smiling*), children again would be expected to reuse the same N, V, N components but must come to realize and learn over repeated exposures that these components are arranged in a different order (NNV), marking the NNV structure as different from a canonical NVN structure. A second principle is that chunking appears to be facilitated by the listener automatically activating from language LTM those stored multiword traces (e.g., NVN, NNV) that are similar to the distributional characteristics of the input (Arnon & Snider, 2010; Arnon et al., 2017; Christiansen & Arnon, 2017). Multiword representations that have an earlier age of acquisition or have been encountered frequently (i.e., canonicals) enjoy faster activation than those that have been acquired at a later age or have been encountered less often such as noncanonicals (Bannard & Matthews, 2008).

Ferreira and colleagues (Ferreira, Bailey, & Ferraro, 2002; Ferreira & Patson, 2007; Karimi & Ferreira, 2016) propose a “good-enough” model of sentence processing. These authors argue that, in most situations, listeners do not create a fully specified syntactic representation of the input (e.g., one that faithfully reflects subject–verb agreement). Rather they generate a shallow or good-enough representation of structure and meaning using a fast and frugal processing heuristic to satisfy the immediate demands of comprehension. Listeners create a representation that is globally consistent with the input, one that tends to be a simple default NVN/agent–action–patient representation. If the initial syntactic representation is incorrect, listeners may rely on the plausibility of the sentence to yield a global representation. The main point here is that listeners may not be attentive to the full syntactic specificity of a sentence to arrive at a generally correct interpretation of the input.

Both the chunk-and-pass and good-enough models are possible workarounds for the difficulties posed on comprehenders by a transient speech signal and limited memory storage. Listeners appear to (a) chunk local information such as articles and nouns quickly and pass this information upward for more complete processing (e.g., creation of clauses) and (b) hold such representations in an active state while new information is processed. Listeners also tend to generate default NVN representations of what they heard. Both perspectives might be right, or they may work in a complementary manner. We will return to these views to help us contextualize our findings.

This Study

Approach and Goals

Our aim was to understand the structural relationship between cognitive processing and syntactic sentence comprehension in young school-age children with and without DLD. Our initial efforts center on modeling syntactic comprehension given the need to develop models from the ground up. To this end, we have begun by modeling children's comprehension of implausible sentences because comprehension is guided by structural cues alone. The goal of model building has centered on determining the direct and/or indirect influences of select cognitive mechanisms on the comprehension of children with and without DLD.

We administered measures of syntactic comprehension and a variety of cognitive measures to large samples of school-age children with DLD and TD children. We chose a set of cognitive processes that mapped hypothesized constructs onto latent variables (sets of observed measures) that we believe are relevant to sentence comprehension and motivated by their theoretical and/or empirical implication to sentence comprehension in school-age children with and without DLD. The mechanisms included FLDR, CATT, cWM, and language LTM. Consistent with the independent clusters basis for creating latent variables (McDonald, 1999), there were two correlated measures for each latent variable. We conducted a confirmatory factor analysis (CFA) to determine whether the cognitive measurement model containing these four constructs, referred to as the GEM (Gillam–Evans–Montgomery) model, fit the cognitive data for all the children combined. A four-factor measurement model comprising the FLD-R, CATT, cWM, and LTM-LK latent variables yielded a very good model fit for all the children combined ($\chi^2 = 31.48$, $df = 21$, $p = .066$, root mean square error of approximation [RMSEA] = .046, 90% CI [0.000, 0.078], comparative fit index [CFI] = .985, standardized root mean square residual [SRMR] = .028) as well as for a multigroup model that accounted for separate TD and DLD group variance ($\chi^2 = 60.99$, $df = 52$, $p = .183$, RMSEA = .038, 90% CI [0.000, 0.073], CFI = .984, SRMR = .049]. We then conducted path analyses using structural equation modeling (SEM) to assess the direct and indirect relationships between the constructs of the CFA measurement model and observed measures of sentence comprehension.

Relationship Between Cognitive Processing and Comprehension: Models Tested and Hypotheses

Five multigroup models of the structural relationship between cognitive processing (FLD-R, CATT, cWM, and language LTM) and the comprehension of canonical and noncanonical structures were tested.

Direct model of effects of each cognitive mechanism on comprehension. A direct model (see Figure 1) in which each of the cognitive mechanisms from the GEM measurement model has a direct influence on comprehension was the first model tested. We did not expect this model to adequately fit the comprehension data given our assumption that cWM should mediate the other mechanisms' influence on comprehension (see below).

Indirect models with FLD-R, CATT, and LTM as separate mediators. The FLD-R mediation model (see Figure 2) was an indirect model in which FLD-R mediated the effects of the covariate predictors on comprehension. The CATT mediation model (see Figure 3) was also an indirect model in which CATT mediated the effects of the other three mechanisms on comprehension. The LTM-LK mediation model (see Figure 4) was an indirect model in which LTM mediated the influences of the cognitive predictors on comprehension. Again, we did not expect any of these models to fit the comprehension data given our assumption about the role of cWM.

Indirect effect model with cWM as a mediator. We predicted that the cWM mediation model (see Figure 5), in which cWM mediated the effects of the other cognitive mechanisms on comprehension, would be the only model to adequately describe the children's comprehension. First, it is cWM that has been shown to relate to FLD-R, CATT, and LTM. Second, the concurrent processing–storage function of cWM is exactly why we think that cWM should play the pivotal mediating role in comprehension because it should (a) allow children to attend to and process the linguistic input as well as toggle between processing the input and storing the products of processing; (b) provide an interface with children's general ability to process/recognize and store novel patterns in the input, especially given the semantic implausibility of our sentences; and (c) provide an interface with children's more specific ability to rapidly chunk the input into meaningful linguistic units based on knowledge of word order patterns and then store and combine the different chunks into a coherent sentence representation, with instances of the miscomprehension of noncanonical structures as NVN structures. On this view, the cWM mediation model aligns well with both the chunk-and-pass and good-enough models of sentence comprehension.

Figure 1. Gillam–Evans–Montgomery structural equation model of the direct relationships between the exogenous latent variables of controlled attention (CATT) indexed by the manifest variables of auditory sustained attention (SusAtt), auditory attention switching (AttSW), fluid reasoning (FLD R) indexed by Leiter-R Figure Ground (Leiter FG), Leiter-R Sequential Order (Leiter SO), Leiter-R Repeated Patterns (Leiter RP), complex working memory (cWM) indexed by verbal working memory (WJ-AWM) and auditory working memory for tones (HI-LoW), and long-term memory language knowledge (LTM-LK) indexed by narrative language comprehension (TNL-REC) and narrative language expression (TNL-EXP); a propensity score (PROP S) control; and the endogenous observed variables canonical sentence comprehension accuracy (CANACC) and noncanonical sentence comprehension accuracy (NOCANACC). Path standardized *YX* estimates for the typically developing and developmental language disorder groups are shown next to arrows. Leiter-R = Leiter International Performance Scale–Revised.

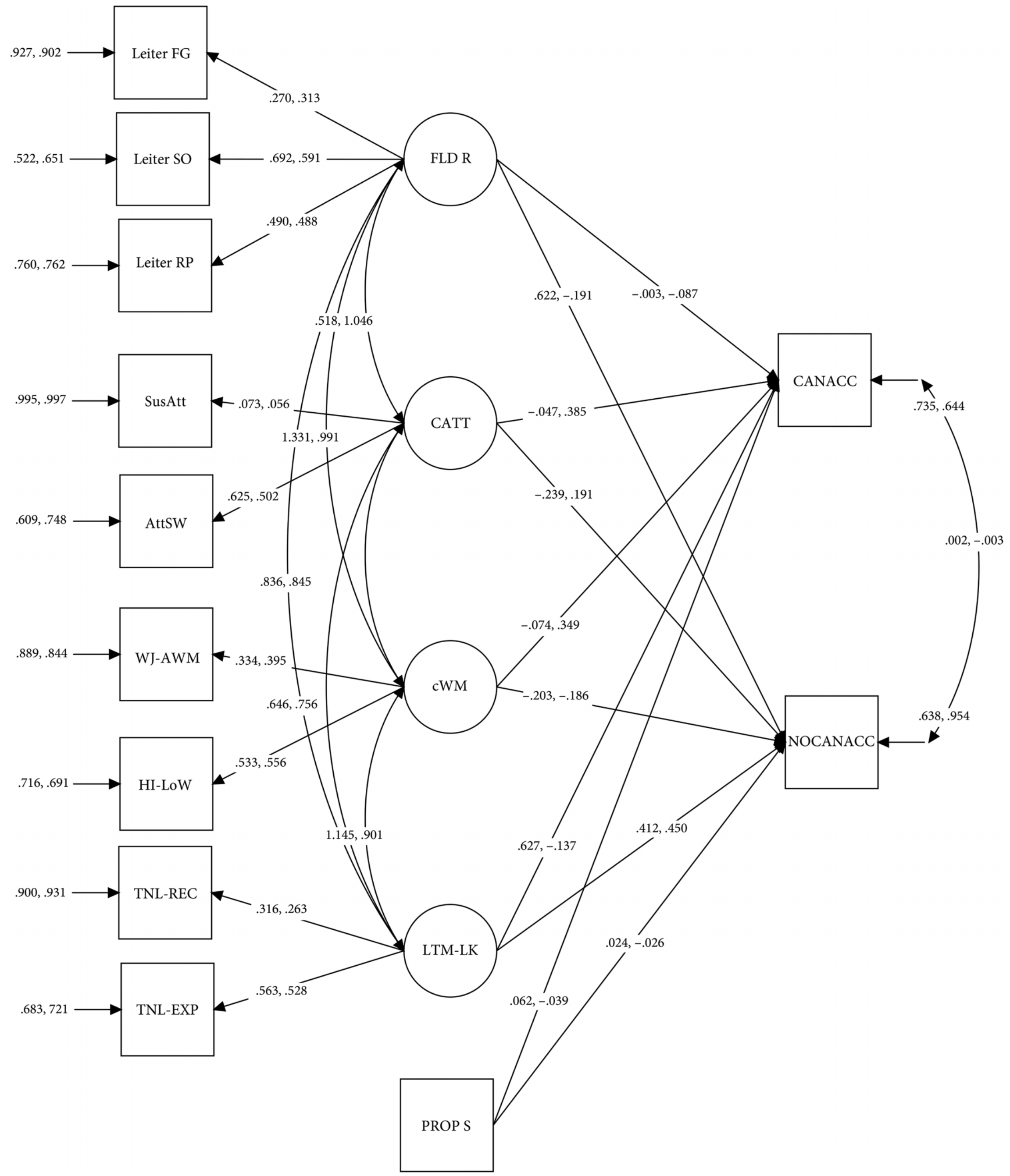
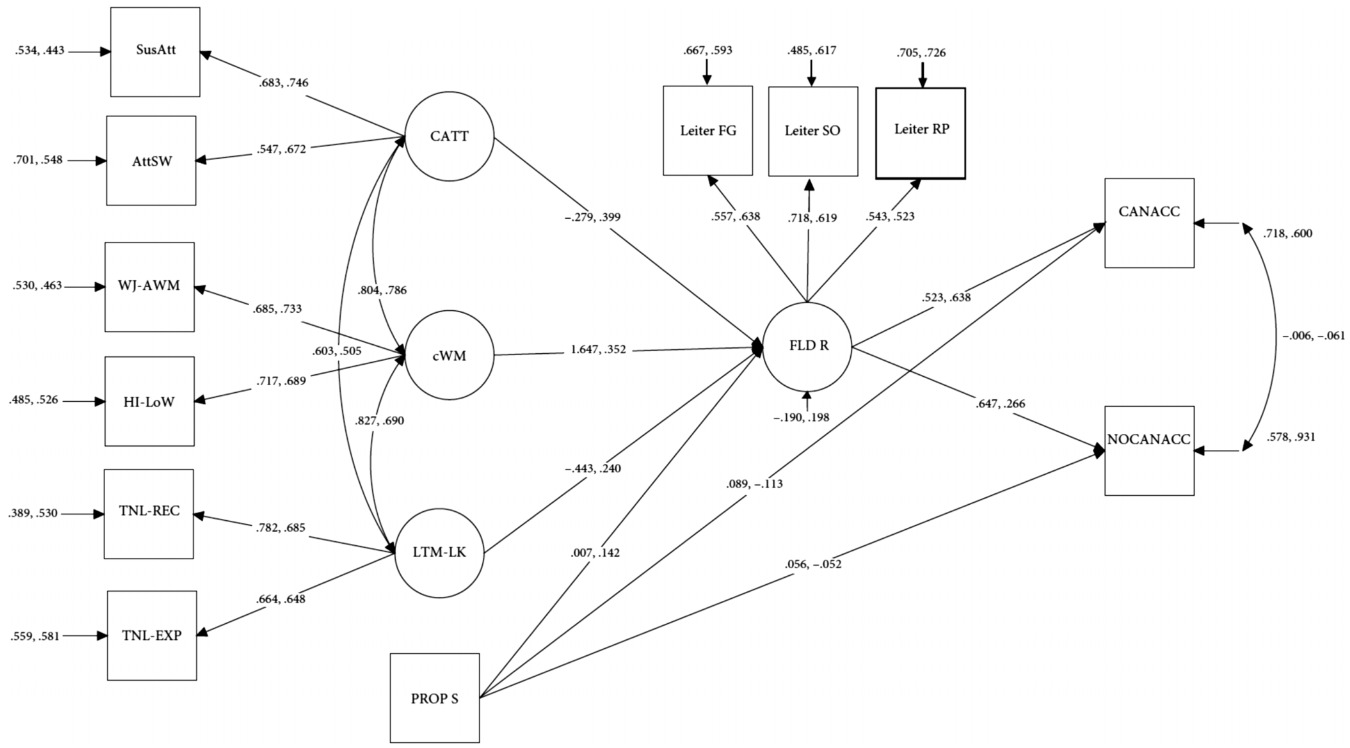


Figure 2. Structural equation model of fluid reasoning (FLD R) mediating the relationships between the exogenous latent variables of controlled attention (CATT), complex working memory (cWM), and long-term memory language knowledge (LTM-LK); a propensity score (PROP S) control; and the endogenous observed variables canonical sentence comprehension accuracy (CANACC) and noncanonical sentence comprehension accuracy (NOCANACC). Path standardized *YX* estimates for the typically developing and developmental language disorder groups are shown next to arrows. Insignificant direct paths from predictor variables to CANACC and NOCANACC are not represented.



Method

Participants

Participants were 234 children between the ages of 7 and 11 years: 117 children with DLD (Age$_M$ = 9;5 [years; months]) and 117 TD children (Age$_M$ = 9;5). Children were recruited from four regions of the United States: Athens, Ohio; Logan, Utah; San Diego, California; and Dallas, Texas. Children were recruited through various school systems, community centers, and university-sponsored summer camps for children.

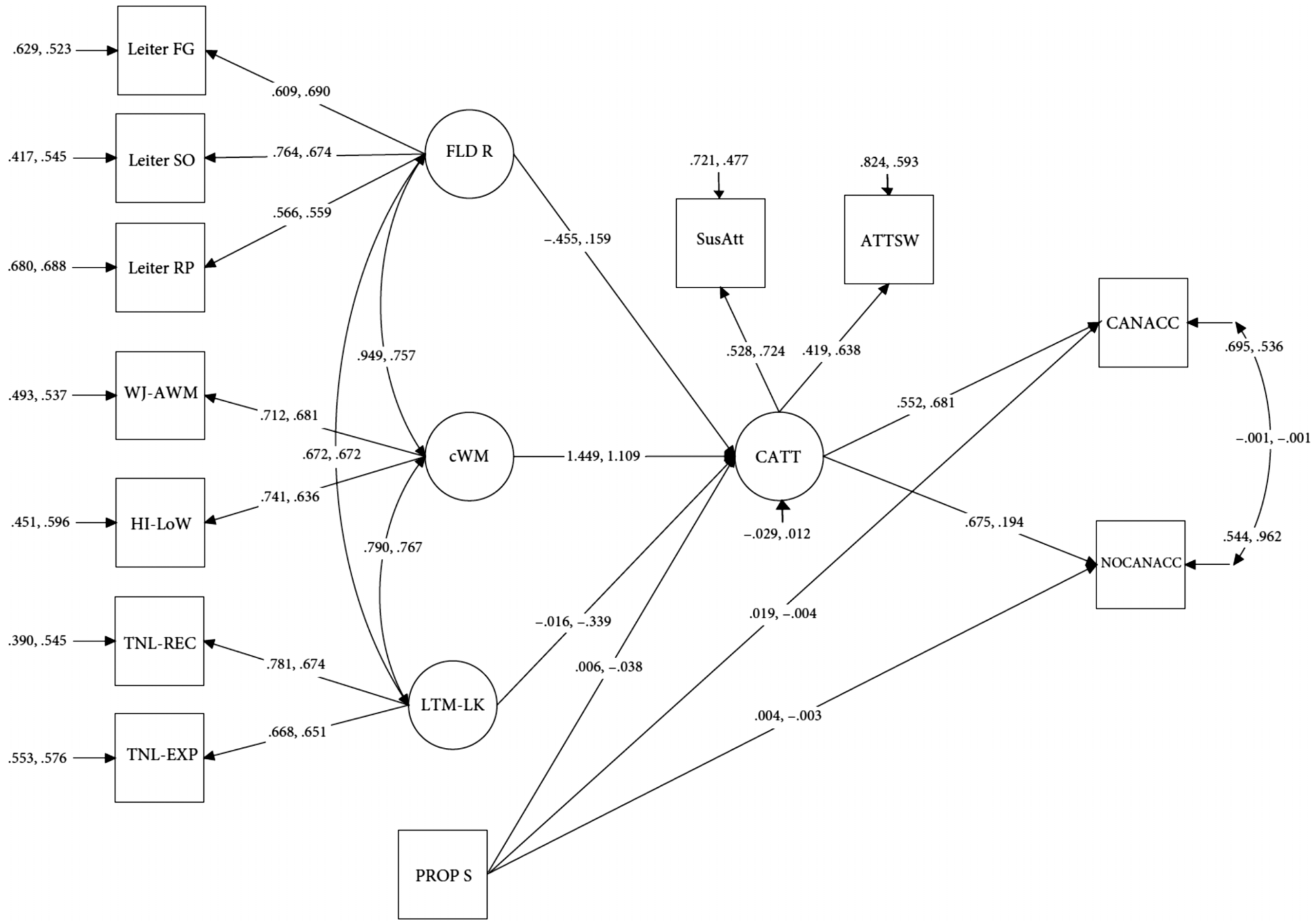
To reduce potential selection bias that might influence the results from the study, it was necessary to develop a standard approach to the definition of participants as DLD or TD and matching of the groups. To define the participants as DLD or TD, we used a composite *z* score (see below). To match the groups to prevent selection bias and control for critical developmental and socioeconomic factors that are known to moderate performance on cognitive tasks, we used a propensity-matching procedure (see below).

The degree of exposure to a second language was controlled, with English being the primary language spoken by the children. Similar to Bedore et al. (2012), parents provided a detailed account of their children's language use at home and in school. Bedore et al. found that measures of English semantics and morphosyntax in a large sample of bilingual kindergartners were not affected until children spoke a second language approximately 80 min each day. Taking a conservative approach, we excluded any child who spoke more than an average of 30 min of another language in the home or at school each day.

Children had typical medical history and no neurological impairment or psychological/emotional disturbance based on parent report. Participants also had (a) typical-range hearing sensitivity bilaterally for the frequencies 500 Hz through 4 kHz (American National Standards Institute, 1997), (b) typical-range articulation on the Articulation subtest of the Test of Language Development–Primary: Fourth Edition (Newcomer & Hammill, 2008), and (c) typical or corrected vision. All participants had FLD-R scores that were in the normal range on the visualization and reasoning battery of the Leiter International Performance Scale–Revised (Leiter-R; Roid & Miller, 1997). Although the children in both the TD and DLD groups exhibited normal-range FLD-R, the children in the TD group obtained a significantly higher Leiter-R score than the children in the DLD group, $F(1, 233) = , p < .0001, \eta^2 = .17$.

Four language measures were used to determine DLD/TD classification. These were the receptive and expressive portions of the Comprehensive Receptive and Expressive

Figure 3. Structural equation model of controlled attention (CATT) mediating the relationships between the exogenous latent variables of fluid reasoning (FLD R), complex working memory (cWM), long-term memory language knowledge (LTM-LK); a propensity score (PROP S) control; and the endogenous observed variables canonical sentence comprehension accuracy (CANACC) and noncanonical sentence comprehension accuracy (NOCANACC). Path standardized *YX* estimates for the typically developing and developmental language disorder groups are shown next to arrows. Insignificant direct paths from predictor variables to CANACC and NOCANACC are not represented.



Vocabulary Test–Second Edition (CREVT-2; Wallace & Hammill, 1994) and the Concepts and Following Directions subtest and Recalling Sentences subtest of the Clinical Evaluation of Language Fundamentals–Fourth Edition (CELF-4; Semel, Wiig, & Secord, 2003). The CREVT-2 (Wallace & Hammill, 1994) is a measure of children's receptive and expressive lexical knowledge, and the two CELF-4 (Semel et al., 2003) subtests are indices of sentence-level receptive and expressive knowledge and abilities. Because two of the subtests were standardized with deviation quotients ($M = 100$, $SD = 15$) and two were standardized with scaled scores ($M = 10$, $SD = 3$), we converted each child's norm-referenced scores for the four subtests to z-score scale ($M = 0$, $SD = 1$) representing the number of standard deviations (*SD*s) from the mean on each subtest. From these z scores, a final mean composite z score was then calculated for each child based on the three lowest of these four z scores.

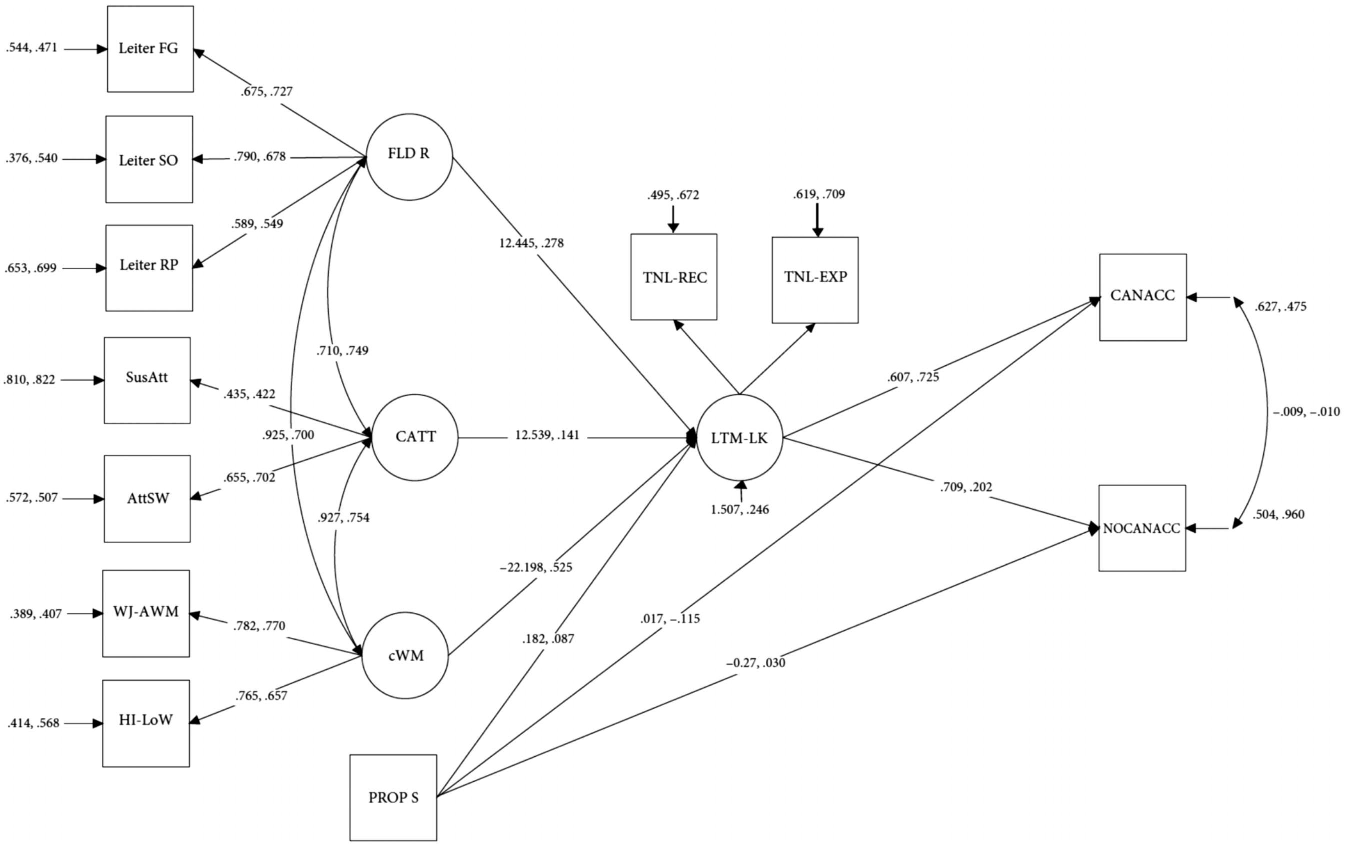
DLD and TD Classification

Children were classified as DLD if their mean composite language z score on their three lowest of the four subtests was at or below −1 *SD,* which is consistent with the Diagnostic and Statistical Manual of Mental Disorders–Fifth Edition (American Psychiatric Association, 2013) definition of language disorder, multidimensional systems for defining DLD (Leonard, 2014; Tager-Flusberg & Cooper, 1999), and other studies (Conti-Ramsden, Ullman, & Lum, 2015; Montgomery et al., 2017). Tomblin, Records, and Zhang (1996) reported that the average language z score for the children identified with the EpiSLI model was −1.14, and approximately 5% of their specific language impairment group had average z scores between −1 and 0. In keeping with the EpiSLI classification,[2] the average composite z score for

---

[2] EpiSLI refers to the epidemiologically derived diagnostic system designed to identify children with DLD (Tomblin et al., 1996).

Figure 4. Structural equation model of long-term memory language knowledge (LTM-LK) mediating the relationships between the exogenous latent variables of fluid reasoning (FLD R), controlled attention (CATT), complex working memory (cWM); a propensity score (PROP S) control; and the endogenous observed variables canonical sentence comprehension accuracy (CANACC) and noncanonical sentence comprehension accuracy (NOCANACC). Path standardized *YX* estimates for the typically developing and developmental language disorder groups are shown next to arrows. Insignificant direct paths from predictor variables to CANACC and NOCANACC are not represented.



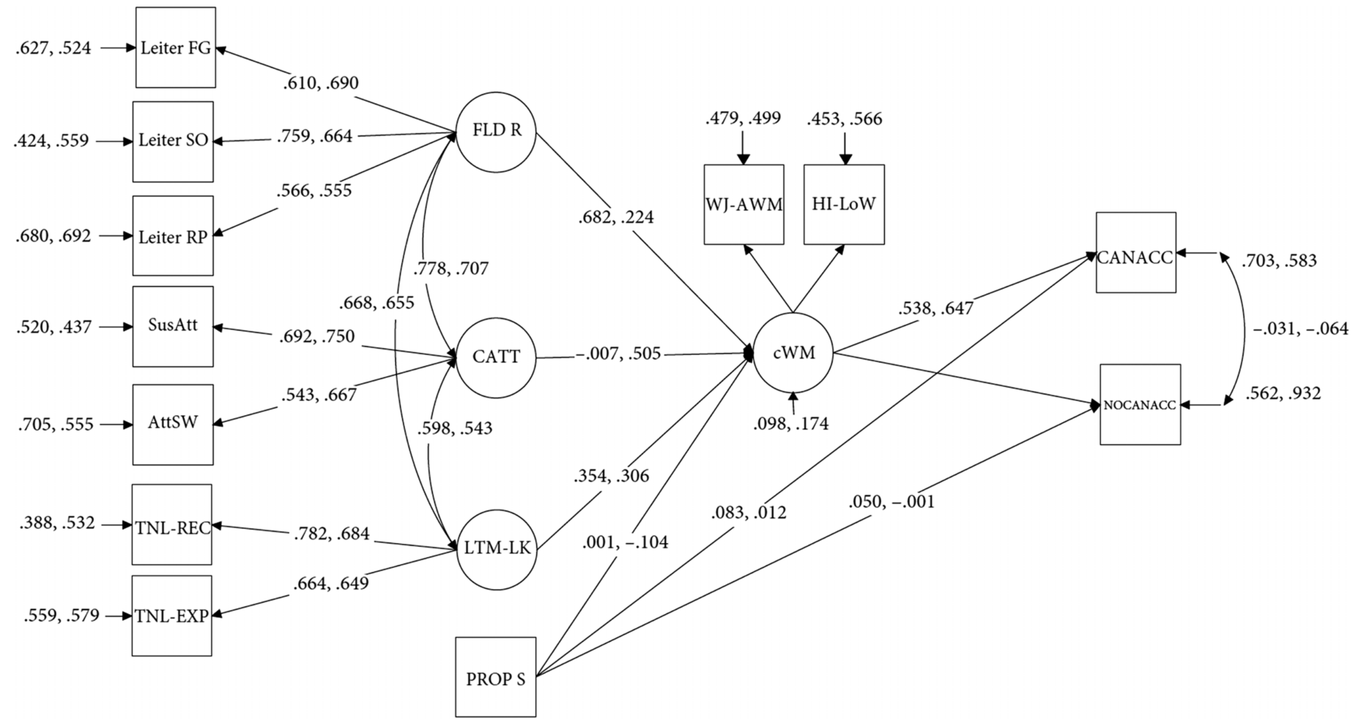
the DLD group in this study was −1.48, with an *SD* of 0.39 (range = −2.73 to −1.00). The overwhelming majority of the children in the DLD group (84.6%) had mixed receptive–expressive disorders. A few children (14.5%) exhibited expressive-only disorders, and just 1% exhibited receptive-only disorders. With respect to the language domain, 74.4% of the children performed at or below the criterion value on subtests in both lexical and sentential domains; 18.8% had difficulties on the grammatical subtests only; and 6.8% had difficulties on the lexical subtests only.

Children were defined as having typical language if their mean composite language *z* score was greater than −1 *SD*. The average composite *z* score for the TD group was 0.08 (*SD* = 0.60, range = −0.96 to 1.89). Relative to the DLD group, the TD group attained a significantly higher mean composite *z* score, $F(1,233) = 556.74$, $p < .0001$, $\eta^2 = .71$. In addition, the TD group achieved a significantly higher score on each of the four language measures: CREVT-2 (Wallace & Hammill, 1994) receptive portion, $F(1, 233) = 61.85$, $p < .0001$, $\eta^2 = .21$; CREVT-2 (Wallace & Hammill, 1994) expressive portion, $F(1, 233) = 37.31$, $p < .0001$, $\eta^2 = .14$; CELF-4 (Semel et al., 2003) Concepts and Following Directions, $F(1, 233) = 50.29$, $p < .0001$, $\eta^2 = .18$; and CELF-4 (Semel et al., 2003) Recalling Sentences, $F(1, 233) = 63.30$, $p < .0001$, $\eta^2 = .21$. Similar to the EpiSLI study, there was minimal overlap in our DLD and TD groups. Entrance test data for both groups appear in Table 1.

Propensity Matching

To avoid selection bias and distortion of the results due to differences in participant enrollment, propensity score matching was used to create the DLD and TD groups from a larger pool of 383 children (127 DLD, 256 TD). Propensity matching is a quasiexperimental approach that approximates the conditions of a randomized experiment by creating control (TD) and experimental (DLD) groups that are balanced on a variety of variables. Propensity scores represent the probability of assignment to either the DLD or TD group (the counterfactual condition) based on a

Figure 5. Gillam–Evans–Montgomery structural equation model of complex working memory (cWM) mediating the relationships between the exogenous latent variables of fluid reasoning (FLD R), controlled attention (CATT), long-term memory language knowledge (LTM-LK); a propensity score (PROP S) control; and the endogenous observed variables canonical sentence comprehension accuracy (CANACC) and noncanonical sentence comprehension accuracy (NOCANACC). Path standardized *YX* estimates for the typically developing and developmental language disorder groups are shown next to arrows. Insignificant direct paths from predictor variables to CANACC and NOCANACC are not represented.



vector of observed covariates.[3] To achieve this sample size, we oversampled TD children by a 2:1 ratio relative to the children with DLD. Using multivariate logistic regression, a single propensity score was calculated for each of the 383 children using the moderating variables of age (continuous variable), gender, mother's education level (no college degree [high school, some college but no degree] vs. college degree [associate, bachelor's, master's, or doctorate]), and family income (annual income < $30,000 vs. annual income > $30,000). Mother's education and family income were used as proxies for socioeconomic status (SES; Shavers, 2007). The nearest-neighbor matching method was then used to match individual children with DLD to a TD counterpart. This procedure yielded 117 DLD–TD multidimensionally matched samples.[4] Subsequent nonparametric analyses indicated the groups were not significantly different with respect to age, gender, mother's education, or family income. Demographic data for the two groups are presented in Table 2.

General Testing Procedure

Children were seen individually in a quiet test room. The standardized testing and experimental tasks were completed over three visits, each lasting about 2.5 hr. The order of the standardized assessments and experimental tasks were counterbalanced across visits and participants. The auditory tasks were presented under noise-canceling headphones at a listening level of 55–75 dB SPL. All of the children successfully completed practice trials prior to moving to the experimental portion of each task. E-Prime.v1 (Schneider, Eschman, & Zuccolott, 2002) was used to deliver most of the tasks and to collect the children's responses.

Cognitive Constructs and Measures

As noted, the cognitive constructs in the GEM model include FLD-R, CATT, cWM, and language LTM (LTM-LK). At least two measures were used to index each of the constructs. Below, we provide a brief description of each

---

[3] A *propensity score* is the conditional probability of a child being enrolled in the DLD or control (TD) group given his or her key baseline characteristics (in our case, age, gender, mother's education, family income). Because of its ability to match groups on a high-dimensional set of characteristics, that is, simultaneous matching on several categorical and continuous variables, propensity score technique has become a critical statistical method in modern clinical research (Rosenbaum & Rubin, 1983; D'Agostino, 1998).

[4] Only 10 of the 127 children with DLD were excluded because of lack of an appropriate TD match.

Table 1. Mean standard scores and standard deviations on the norm-referenced test measures administered to the children with developmental language disorder (DLD) and typically developing (TD) children.

Measures	DLD (*n* = 117)	TD (*n* = 117)	Cohen's *d*
Fluid reasoning			
Leiter-R[a]			
M	98	110	−0.77
SD	13	14	
Range	76–139	76–141	
Lexical			
CREVT-R[b]			
M	87	105	−1.22
SD	9	11	
Range	62–112	81–146	
CREVT-E[c]			
M	81	101	−1.32
SD	10	12	
Range	54–101	69–134	
Sentential			
CELF-4 Concepts & Directions[d]			
M	6	11	−1.33
SD	3	2	
Range	1–13	6–15	
CELF-4 Recalling Sentences[e]			
M	5	10	−1.51
SD	2	2	
Range	1–11	4–18	
Qualifying *z* score[f]			
M	−1.49	0.08	−3.10
SD	0.39	0.60	
Range	−2.73 to −1.0	−0.96 to 1.89	

[a]Leiter-R: Average standard score on four nonverbal subtests (Figure Ground, Form Completion, Sequential Order, and Repeated Patterns) from the Visualization and Reasoning Battery of the Leiter International Performance Scale–Revised (*M* = 100, *SD* = 15). [b]CREVT-R: Comprehensive Receptive and Expressive Vocabulary Test, receptive portion (*M* = 100, *SD* = 15). [c]CREVT-E: Comprehensive Receptive and Expressive Vocabulary Test–Second Edition, expressive portion (*M* = 100, *SD* = 15). [d]CELF-4 Concepts & Directions: Clinical Evaluation of Language Fundamentals–Fourth Edition, Concepts and Following Directions subtest (*M* = 10, *SD* = 3). [e]CELF-4 Recalling Sentences: Clinical Evaluation of Language Fundamentals–Fourth Edition, Recalling Sentences subtest (*M* = 10, *SD* = 3). [f]Qualifying *z* score: Average *z* score on the three lowest lexical and sentential measures.

measure. A fuller description and information about reliability for each task appear in Supplemental Material S1.

FLD-R

Three measures from the Leiter-R were used to form the latent variable representing fluid/analytic reasoning—figure ground, sequential order, and repeated patterns. The primary executive abilities of interest were the ability to recognize patterns, to reason, and to solve novel problems independent of prior knowledge. Recall that we used FLD-R not as a categorization measure but as a descriptive measure. The dependent variable was the total number of items correct across the three subtests.

CATT

Sustained auditory attention and auditory attention switching represented the construct of CATT. In the sustained task, children responded to a target (number sequence 1–9) while withholding a response to nontargets as they listened to a stream of random digits. The primary dependent variable was PR, a discrimination index representing sensitivity to responding to target items while withholding a response to nontargets (PR = H − FA where H is hits and FA is false alarms). In the switching task, children were told that they would hear a beep in one ear or the other and to listen only to the speaker in that ear (target ear) while ignoring the other ear. They were told that, after a short period, they would hear a beep in the other ear and then switch to listening to that ear. Each speaker (male, female) said either a letter (A–E) or a number (1–5). Children were instructed to touch a group of letters or a group of numbers on the monitor, depending on what the speaker was saying in the target ear. The primary dependent variable was percent correct on switch trials.

cWM

Children completed two tasks involving concurrent processing and storage. Importantly, neither measure's processing component involved sentence comprehension. One task invited the children to recall in serial order words

Table 2. Participant demographics for the children with developmental language disorder (DLD) and typically developing (TD) children.

Demographic	DLD (*n* = 117)	TD (*n* = 117)
Age$_M$ (years;months)	9;6	9;6
Gender		
Male	57%	63%
Female	43%	36%
Race and ethnicity		
White (not Hispanic)	61%	72%
African American	10%	0%
Hispanic	12%	12%
Asian	4%	4%
American Indian, Native Hawaiian	3%	3%
More than one race	10%	9%
Mother's education		
No response	1%	1%
High school degree	20%	16%
Some college	30%	27%
Associate degree	17%	11%
Bachelor's degree	24%	23%
Graduate degree	6%	20%
Family income (dollars)		
0–25,000	42%	32%
26,000–50,000	21%	22%
51,000–75,000	16%	15%
> 75,000	21%	31%

followed by digits in serial order from a randomly presented list of items. The other required children to keep count and report the number of high and low tones that were presented. The advantage of these tasks over conventional listening span tasks is that any influence of cWM on sentence comprehension that we might observe in this study cannot be attributed to a shared comprehension component between the cWM and sentence comprehension tasks (Ahmad Rusli & Montgomery, 2017). The dependent variable was total trials correct and percent trials correct in Task 1 and Task 2, respectively.

LTM-LK

The two measures of language LTM were the comprehension and production portions of the Test of Narrative Language (Gillam & Pearson, 2004). The Test of Narrative Language is an overall index of lexical–semantic and grammatical aspects of declarative and procedural memory. Gillam and Johnston (1992) argued that narrative comprehension and production involve organizing verbal material into meaningful sequences of propositions at the sentence level and narrative level. Prior to constructing narrative structure, however, children must first assemble words into syntactically–semantically meaningful sentences using their lexical and syntactic–semantic knowledge. The dependent variable was total raw score combining correct items on the receptive and expressive portions of the test.

Sentence Comprehension

Children's ability to comprehend who did what to whom was assessed using our "Whatdunit?" agent selection task (Montgomery et al., 2017). Children listened to implausible sentences and selected the agent of the sentence by touching the image of the agent noun from an array of three images presented immediately after the sentence. Construction of the sentences was guided by numerous constraints (see Montgomery et al., 2017, for details). The central design feature to note here is that all of the semantic and real-world/pragmatic cues were removed from the sentences rendering them implausible (i.e., expressing highly improbable events), thus forcing the children to depend on structural (word order) cues to determine which NP was the agent. Implausibility was created by using only inanimate nouns and violating typical predicate–argument structure.

Stimuli

The canonical structures included 33 SVOs (*The square had changed the bed under the very new dry key*) and 33 center-embedded subject relatives in which the NP1 functioned as the subject in the relative clause (*The watch that had hugged the truck behind the kite was bright*). The noncanonical structures included 33 verbal *be* passives (*The watch was bumped by the wheel near the very bright clock*) and 33 object relatives in which NP1 functioned as object/patient in the relative clause (*The chair that the bread had splashed under the square was new*). Each sentence included a third NP appearing in a prepositional phrase near the end of the sentence for controlling the length of the sentences across sentence type. The prepositional phrase was not critical to sentence comprehension. The task has strong validity and reliability as described in detail in Montgomery, Evans, Gillam, Sergeev, and Finney (2016). The internal construct validity values were .84 for the two canonical structures and .89 for the two noncanonical structures, supporting the distinction between canonical and noncanonical constructions.

Procedure

Children were told that they would hear a man saying some funny-sounding sentences and then see three pictures after the sentence on the monitor. They were told to touch the object picture that performed the action. The primary dependent variable was comprehension accuracy (correct agent selection).

Data Analysis

SEM

We used SEM techniques to test the relationships among the four cognitive constructs in the GEM model and our measures of canonical and noncanonical sentence comprehension. We used the multigroup function in Mplus V.7 (Muthén, Muthén, & Asparouhov, 2016), which accounted for potential differences in variance between the TD and DLD groups. This procedure enabled us to estimate the path coefficients among the covariates of interest for the DLD and TD groups separately while controlling for the factors comprising the propensity score. We modeled the data using the maximum likelihood robust estimator and grand-mean centering of observed scores. All loadings were freely estimated. This approach to modeling was robust to violations of model assumptions related to heteroscedasticity and nonnormal distributions.

In the modeling literature, the goodness-of-fit of models is most often evaluated by a combination of indices including the model chi-square test, Browne and Cudeck's (1993) RMSEA, Bentler's (1990) CFI, and the SRMR. The criterion for an acceptable fit varies for each index. The model chi-square test assesses the magnitude of discrepancy (e.g., the badness of fit) between the sample and fitted covariance matrices (Hu and Bentler, 1999), with an insignificant result indicating a good model fit (Barrett, 2007). Because the chi-square statistic is sensitive to sample size, it often rejects the model when large samples are used (Bentler & Bonnet, 1980), as in the case of our investigation. When comparing models of large samples, the model with the smallest chi-square value is considered the best. The RMSEA (Steiger, 1990) indicates the extent to which the model fits the population's covariance matrix. An RMSEA of less than .07 indicates an acceptable fit, and an RMSEA of .05 or less indicates a close fit of the model in relation to the degrees of freedom (Browne & Cudeck, 1993). The CFI compares the chi-square value of the model to the null model (worst case scenario) in a manner that accounts for sample size. CFI values at or above .95 indicate a good model fit (Hu & Bentler, 1999), with values above .97 indicating a very good fit. The SRMR indicates how well sample variances, covariances, and means were reproduced by the model. Simulation studies suggest that a CFI greater than .95 in combination with an SRMR less than .08 are reasonable cutoff values for good fitting models (Hu & Bentler, 1999).

Results

Preliminary Analyses

A preliminary multivariate analysis of covariance (using total raw score on the nonverbal IQ test as the covariate) revealed that the TD group earned significantly higher scores than the DLD group across all cognitive processing measures comprising the cognitive constructs ($p < .001$), with the exception of the sustained attention measure, which yielded a significance value of .018. The very small significance values are a function of the large sample size and are minimally informative. We calculated Cohen's d standardized mean difference scores to better represent the extent of group differences on the cognitive measures. The group differences for our FLD-R tasks were all in the moderately large range (Leiter-R Figure Ground $d = -0.53$, Leiter-R Sequential Order $d = -0.64$, and Leiter-R Repeated Patterns $d = -0.56$), whereas the group differences for the CATT tasks were in the moderate to moderately large range (Sustained Attention $d = -0.36$, Attention Switching $d = -0.54$). There were large to very large group differences ($d = -0.95$ to -1.08) for the cWM and LTM-LK tasks. Descriptive data for the cognitive measures for each group appear in Supplemental Material S2.

Bivariate correlations were computed among all the experimental measures. The majority of the measures significantly correlated in the moderate to moderately large range ($r = .35–.55$, $p \leq .01$). As expected, the highest correlations ($r > .50$) tended to occur for measures comprising each of the four latent variables (FLD-R, CATT, cWM, LTM-LK) as opposed to measures across the latent variables, providing additional support for our latent variables. Correlation data appear in Supplemental Material S3. Recall that the results of a CFA indicated a very good model fit for the measurement model comprising the constructs of FLD-R, CATT, cWM, and LTM-LK when the group variance was combined and when it was analyzed separately.

SEM: Determining the Structural Relationship Between Cognitive Processing and Sentence Comprehension

Using SEM, we first tested the hypothesis that the four latent cognitive variables in the GEM model were directly related to the sentence comprehension measures for the two groups (see Table 3 for all model fit statistics for each model). Figure 1 (direct model) is a diagram representing a direct covariate model in which performance on the canonical and noncanonical sentence comprehension tasks was regressed on the four cognitive variables (FLD-R, CATT, cWM, and LTM-LK) and the propensity score (which simultaneously controlled for age, gender, mother's education, and family income). There was an acceptable model fit, but the model was not positive definite (revealing multicollinearity problems), and none of the direct paths from the latent cognitive variables or the propensity score variable to canonical or noncanonical sentence comprehension were significantly different from zero for either the TD or DLD group. Multicollinearity occurs when the standardized YX slopes are greater than 1 as a result of two or more highly correlated covariates. Multicollinearity is problematic because it leads to standard errors of parameter estimates that are inappropriately large and estimated coefficients that are overly sensitive to small changes in the data (Muthén et al., 2016).

Because the direct effects model was not positive definite and none of the cognitive variables significantly influenced sentence comprehension when the other three latent variables and propensity score were held constant, we respecified the model as four indirect structural equation models (see Figures 2–5), each with one of the four latent cognitive variables as a mediator. Each of these models depicts three covarying latent variables that indirectly affect canonical and noncanonical sentence comprehension through a single mediating variable. As with the direct effects model mentioned above, we included the propensity score as a covariate in each model to simultaneously control for the potential moderators of age, gender, mother's education, and family income. This procedure enabled us to partition the total effect of the four cognitive predictors into direct and indirect effects on canonical and noncanonical sentence

Table 3. Model fit statistics for the path analysis model predicting canonical and noncanonical sentence comprehension using the GEM cognitive predictors.

Model	Multigroup fit statistics
Model 4: Covariate direct effects	χ^2 = 107.40, *df* = 94, *p* = .94, RMSEA = .035, 90% CI [0.00, 0.063], CFI = .980, SRMR = .060; NPD
Model 5: FLD-R as mediator	χ^2 = 117.46, *df* = 100, *p* = .93, RMSEA = .039, 90% CI [0.00, 0.064], CFI = .974, SRMR = .056; NPD
Model 6: CATT as mediator	Failed to converge
Model 7: LTM-LK as mediator	Failed to converge
Model 8: cWM as mediator	χ^2 = 109.56, *df* = 100, *p* = .94, RMSEA = .029, 90% CI [0.00, 0.058], CFI = .986, SRMR = .055

Note. GEM = Gillam–Evans–Montgomery model; NPD = not positive definite; FLD-R = fluid reasoning; CATT = controlled attention; cWM = complex working memory; LTM-LK = long-term memory language knowledge; RMSEA = root mean square error of approximation; CFI = comparative fit index; SRMR = standardized root mean square residual.

comprehension while controlling for age, gender, and SES. We left the direct paths between the covarying latent variables and the two observed dependent variables out of the mediation models because the test of the direct effects model revealed that none of the direct effects between the observed sentence comprehension measures and the four latent cognitive variables significantly differed from zero.

Three of the four mediation models proved to be problematic. The FLD-R mediation model (see Figure 2) tested whether FLD-R mediated the influence of CATT, cWM, and LTM-LK on canonical and noncanonical sentence comprehension. This model yielded an acceptable fit but was not positive definite due to multicollinearity problems for the TD group. The CATT mediation model (see Figure 3) evaluated the extent to which FLD-R, cWM, and LTM-LK influenced canonical and noncanonical sentence comprehension indirectly through CATT (the mediator). This model, too, did not converge, as did the LTM-LK mediation model (see Figure 4) in which LTM-LK mediated the relationships between FLD-R, CATT, and cWM and the comprehension of canonical and noncanonical sentences. As noted by Kline (2016), initial estimates of overall model fit are derived by the program and then modified in succeeding calculation cycles. When the estimates from step to step do not settle on a stable value, the process may fail, resulting in a model that does not converge. We increased the number of iterations and tried different starting values, with no success. Therefore, we decided to modify the model with a different mediator.

Tests of the cWM mediation model (see Figure 5) revealed good model fit statistics. Comparing models, the model fit statistics for the cWM mediation model were better than the model fit statistics for the direct model (see Table 3). Therefore, one requirement of mediation, that of better model fit statistics for a mediation model than a direct model, was met. The chi-square values for the cWM mediation model and the direct model were not significantly different according to the Satorra–Bentler scaled chi-square test (TRd = 1.89, *df* = 6), but the RMSEA value for the cWM mediation model was lower than that of the direct model (.029 vs. .035), the CFI value was higher (.986 vs. .980), and the SRMR value was lower (.055 vs. .060). In addition, unlike the direct model, the cWM mediation model did not have multicollinearity problems. Recall that in the direct model, none of the direct effects of FLD-R, CATT, cWM, or LTM-LK were significantly related to canonical or noncanonical sentence comprehension for either the TD or DLD group. However, in the cWM mediation model, the direct effects from cWM (the mediator) to canonical and noncanonical comprehension were significant for each group (see Table 3). Therefore, another requirement for mediation, a statistically significant indirect effect together with a nonsignificant direct effect (Muthén et al., 2016), was also met. In addition, at least one of the indirect effects from a cognitive predictor to canonical and noncanonical comprehension was significant for each group. The indirect effects for FLD-R and LTM-LK to canonical and noncanonical comprehension via cWM were significant for the TD group. For the DLD group, the indirect effects for CATT and LTM-LK to canonical comprehension were significant, but only the indirect effect from CATT was significant for noncanonical comprehension. Thus, cWM mediated the relationships between LTM-LK and canonical comprehension for both groups. This suggests that FLD-R, CATT, and LTM-LK do not affect sentence comprehension independent of WM effect.

It is possible that group differences were not important for the model. That is, treating our language and cognitive processing measures as continuous variables might better represent the relationship between cognition and sentence comprehension, especially if children in our DLD group simply reflected the lowest 10% of the normal distribution. To address this possibility, we recalculated the SEM with cWM as the mediator using the combined data from both groups. Essentially, by omitting the multigroup function in Mplus, all the predictor variables were treated as continuous, and the model fit statistics assessed the covariance structure without accounting for group variance. The model fit was acceptable (χ^2 = 87.57, *df* = 94, *p* = .001, RMSEA = .064, 90% CI [0.043, 0.083], CFI = .954, SRMR = .049) but was not as good as the model fit when group variance was included in the model (χ^2 = 109.56, *df* = 100, *p* = .94, RMSEA = .029, 90% CI [0.00, 0.058], CFI = .986, SRMR = .055). Clearly, the model fit was better when it accounted for the variance that was attributable to each group. Therefore, we looked carefully for potential group differences in the effect sizes (standardized *YX* estimates) of the paths in the model.

The standardized *YX* estimates (see Table 4) can be interpreted as the change in *SD* units in a dependent variable when a predictor variable changes 1 *SD*. For the TD group, a 1 *SD* increase in FLD-R increased canonical and noncanonical sentence comprehension indirectly via cWM (the mediator) by .367 of an *SD* and by .450 of an *SD,* respectively. These are both moderate effect sizes. These values translate to an indirect effect of FLD-R that was about 22% larger for noncanonical sentence comprehension than for canonical comprehension. For the DLD group, a 1 *SD* increase in FLD-R increased canonical sentence comprehension indirectly via cWM by just .145 of an *SD* and only .059 of an *SD* for noncanonical sentence comprehension. Both of these effects were small and nonsignificant. Interestingly, the indirect relationship between FLD-R and canonical sentence comprehension was 153% greater for the TD group than the DLD group. For noncanonical sentences, the indirect influence of FLD-R was approximately 662% greater for the TD group than the DLD group.

We now turn to a closer inspection of the indirect relationship between CATT and sentence comprehension via cWM. For the TD group, a 1 *SD* increase in CATT decreased canonical sentence comprehension by .004 of an *SD* and by .005 of an *SD* for noncanonical comprehension. Neither indirect effect was significant, suggesting that CATT played no detectable role in sentence comprehension in these children. For the DLD group, however, a 1 *SD* increase in CATT increased canonical sentence comprehension indirectly via cWM by approximately .326 of an *SD* and for noncanonical sentences by approximately .132 of an *SD*. Both effects were significant, indicating that CATT played a small but statistically reliable role in the DLD group's sentence comprehension when it was mediated by cWM. Within the DLD group, the indirect influence of CATT through cWM was 124% larger than FLD-R for both canonical and noncanonical sentence comprehension.

Finally, the indirect relationship between LTM-LK and sentence comprehension via cWM is considered. For the TD group, a 1 *SD* increase in LTM-LK increased canonical sentence comprehension by .190 of an *SD* and by .233 of an *SD* for noncanonical comprehension. Both indirect effects were significant. LTM-LK played a moderate role in sentence comprehension in the TD children that was about 23% greater for noncanonical sentences than for canonical sentences. For the DLD group, a 1 *SD* increase in LTM-LK increased canonical sentence comprehension indirectly via cWM by approximately .198 of an *SD* (a significant effect) but only by approximately .080 of an *SD* for noncanonical sentences (an insignificant effect). LTM-LK played a relatively small role in these children's sentence comprehension when it was mediated by cWM. The indirect relationship between LTM-LK and canonical sentence comprehension was nearly the same for both groups. However, for the noncanonical sentences, the indirect effect of LTM-LK was 191% greater for the TD group compared with the DLD group.

Discussion

The goal of this study was to determine the contribution of select cognitive mechanisms we believed to be relevant to children's sentence comprehension. Because these cognitive mechanisms were not directly observable, we modeled them as latent cognitive constructs in a measurement model, with each construct indexed by two or more

Table 4. Standardized model results (standardized *YX* values) for critical direct and indirect paths in Model 5 for the children in the typically developing (TD) and developmental language disorder (DLD) groups.

	TD group			**DLD group**		
Critical paths	**Estimate**	***SE***	***p***	**Estimate**	***SE***	***p***
cWM						
Direct from fluid reasoning to cWM	.682	.286	.017*	.224	.205	.275
Direct from attention to cWM	−.007	.270	.979	.505	.189	.008**
Direct from LLTM to cWM	.354	.152	.020	.306	.153	.046*
Direct from propensity score to cWM	.001	.072	.986	−.104	.090	.248
Canonical comprehension						
Direct from cWM to canonical comprehension	.538	.077	.001**	.647	.074	.0001**
Direct from propensity score to canonical comprehension	.083	.091	.364	.012	.087	.887
Indirect from fluid reasoning to canonical comprehension via cWM	.367	.158	.020*	.145	.133	.277
Indirect from attention to canonical comprehension via cWM	−.004	.145	.979	.326	.133	.014*
Indirect from LLTM to canonical comprehension via cWM	.190	.085	.025*	.198	.099	.047*
Noncanonical comprehension						
Direct from cWM to noncanonical comprehension	.660	.071	.0001**	.261	.115	.023*
Direct from propensity score to noncanonical comprehension	.050	.077	.514	−.001	.096	.992
Indirect from fluid reasoning to noncanonical comprehension via cWM	.450	.196	.022*	.059	.062	.345
Indirect from attention to nonanonical comprehension via cWM	−.005	.178	.979	.132	.066	.045*
Indirect from LLTM to nonanonical comprehension via cWM	.223	.103	.024*	.080	.056	.154

Note. cWM= complex working memory; LLTM = language long-term memory.

*$p < .05$. **$p < .01$.

behavioral measures. The use of the four cognitive constructs as independent predictors (i.e., GEM model) of sentence comprehension was supported by a CFA. We then used an SEM approach to assess the direct and indirect relationships between the latent variables in the GEM measurement model and our observed measures of sentence comprehension while controlling for measurement error. Mediation models are often used to explain how or why select predictor variables are related to dependent/endogenous variables, which enabled us to determine the nature of the structural relationship between cognitive processing and sentence comprehension. Tests of our SEM models allowed us to determine the degree to which a cognitive mechanism had an indirect effect on comprehension through another cognitive mechanism (mediator) while simultaneously controlling for variance related to age, gender, mother's education, and family income.

The Relationship Between Cognitive Processing and Sentence Comprehension

We first tested a direct effects model (see Figure 1) depicting the direct influences of FLD-R, CATT, cWM, and LTM-LK on the children's sentence comprehension. This model proved to have a relatively poor fit with the data. None of the latent variables directly contributed to canonical or noncanonical comprehension when the other three cognitive variables were considered as covariates. This pattern held true for both the DLD and TD groups, suggesting that an indirect model, with one latent variable as a mediator, might be a better representation of sentence comprehension for TD and DLD children.

We then tested four mediation models using each of the four cognitive mechanisms as a mediator through which the other three latent variables might influence comprehension. Only the cWM mediation model (see Figure 5) in which cWM mediated the influence of FLD-R, CATT, and LTM-LK on sentence comprehension (while controlling for propensity score) fit the data well. This means that WM functions as the conduit through which the causal effects of FLD-R, CATT, and LTM-LK operate. We used the children's propensity scores as a control. Therefore, the relationships between the cognitive predictors in the GEM model and our sentence comprehension measures were not due to variations on age, gender, or SES. Finally, we found that the model fit statistics for the multigroup model were better than the fit statistics when we combined all the children together. It appears that the model in which cWM played a mediating role was a good characterization of the cognitive contributions to sentence comprehension when we controlled for the variance that was attributable to each group separately. Given that finding, it is not surprising that there were crucial differences between the groups with regard to the magnitude of the indirect influence of FLD-R, CATT, and LTM-LK. The relationship between cognitive processing and sentence comprehension is neither simple nor straightforward but, instead, complex and nuanced.

TD Group

We begin with an explanation of the relationship between cognitive processing and sentence comprehension in the TD children as a way to better understand and contextualize the relationship in the children with DLD. Both FLD-R and LTM-LK, but not CATT, indirectly influenced the TD children's comprehension of canonical and noncanonical sentences through cWM. In this case, stronger FLD-R and LTM-LK abilities led to higher sentence comprehension as a result of cWM abilities. We will discuss what it means for FLD-R and LTM-LK to participate in comprehension before considering the mediating role of cWM on each of these mechanisms.

The role of FLD-R. As noted earlier, FLD-R refers to a general analytic ability to recognize and reason about patterns (Haavisto & Lehto, 2005). In the adult comprehension literature, it has been argued that FLD-R and syntactic processing are to some degree similar as both entail recognizing and interpreting patterns in the input (Andrews et al., 2017). Evidence indicates that FLD-R predicts noncanonical sentence comprehension (Van Dyke, Johns, & Kukona, 2014). In addition, Andrews et al. (2017) found that FLD-R was associated with the comprehension of both noncanonical and canonical sentences. The Andrews et al. study is especially relevant to the present investigation because these authors examined the association between FLD-R and the comprehension of both noncanonical and canonical sentences whose plausibility was either strong (*The comedian that the millionaire funded left town*) or neutral (*The woman that the man helped sang well*). Plausible sentences are easier to comprehend than implausible sentences, whether they are noncanonical or canonical, because listeners can make use of semantic–pragmatic cues as well as structural cues to guide comprehension. As syntactic complexity increases, listeners rely more on plausibility than structure for comprehension. Andrews et al. found that even though FLD-R was significantly related to both canonical and noncanonical sentence comprehension, it accounted for more variance in canonical sentence comprehension than noncanonical sentence comprehension.

We similarly might contend that FLD-R facilitated the children's performance on our unique comprehension task. The children's FLD-R abilities apparently facilitated their recognition of different patterns in the input, enabling them to construct both canonical and noncanonical sentence representations. Interestingly, the influence of FLD-R was 22% greater in the comprehension of noncanonical sentences than canonical sentences, the opposite of the pattern reported by Andrews et al. (2017). Because our sentences included only structural cues, the children had no choice but to rely on word order, which may be why general pattern recognition abilities contributed more strongly to the comprehension of the more difficult noncanonical sentences than canonical sentences. It may also be that we picked up on a developmental issue. Even though these same TD children showed developmental improvement in noncanonical sentence comprehension, the older children's performance was still nowhere near that of adults (60% vs. 87%; Montgomery et al., 2017).

The role of LTM-LK. We earlier suggested that the chunk-and-pass model by Christiansen and associates (Arnon et al., 2017; Chater et al., 2016; Christiansen & Chater, 2016; McCauley & Christiansen, 2013) and the good-enough processing model of Ferreira and colleagues (Ferreira et al., 2002; Karimi & Ferreira, 2016) may be helpful to understand how the children solved our comprehension task. The two most relevant features from the chunk-and-pass model for us are (a) automatic chunking of input into multiword units and (b) activation of such chunks from LTM-LK given their similarity to the distributional properties of the input. The most relevant aspect of the good-enough model is that listeners rely heavily on a fast and frugal processing heuristic for comprehension, one that tends to output an NVN/agent–action–patient representation. This study was not set up to directly assess either of these models, but our results appear to be somewhat consistent with both. First, both groups of children comprehended canonical sentences with significantly greater accuracy than noncanonical sentences, suggesting that the children's NVN canonical templates are more strongly represented in memory and thus more readily retrievable than noncanonical word order templates. Second, consistent with a fast and frugal perspective (Ferreira et al., 2002; Ferreira & Patson, 2007), the most frequent error the children made when misunderstanding a noncanonical sentence was the selection of the patient as the agent, not the noun in the prepositional phrase near the end of the sentence (Montgomery et al., 2017). These findings imply that children created a default NVN representation from noncanonical input. Third, the children's LTM-LK exerted a significant indirect influence on the comprehension of both canonical and noncanonical sentences (Christiansen & Chater, 2016; Cornish et al., 2017; McCauley et al., 2017). Because our cWM measures included no sentence-level comprehension component of any sort, as is the case in listening span measures, we offer strong evidence of a "pure" influence of the basic dual functions of verbal cWM on sentence comprehension—the ability to engage in concurrent verbal processing and storage (see Ahmad Rusli & Montgomery, 2017).

That our sentences were implausible, yet perfectly grammatical, strengthens the argument that the children were able to chunk the input into relevant multiword units and activate these units from LTM-LK to generate sentence representations. Had the children not chunked the input, we would otherwise have to assume they processed the input on a word-by-word basis, which would have likely led to a significantly greater memory load and poorer comprehension of both sentence types, especially noncanonicals. But the children performed with 81% accuracy on the canonical sentences and 52% accuracy (well above chance) on the noncanonicals (Montgomery et al., 2017). Interestingly, LTM-LK had a 23% greater indirect influence on noncanonical sentences than canonical sentences, suggesting that the children relied more heavily on LTM-LK when faced with a sentence pattern in which NP1 was not the agent. This finding would seem to be consistent with the idea that extra language chunking steps were necessary when the structure did not match the first fast and frugal chunk or when word order cues indicate that an "NP1 as agent" assumption could not be met.

The cWM-mediated role of FLD-R. cWM mediated the influence of FLD-R on the children's comprehension. Evidence suggests a moderate to strong relationship between FLD-R and cWM in adults (Burgess, Gray, Conway, & Braver, 2011; Fukuda, Vogel, Mayr, & Awh, 2010; Kane et al., 2004) and TD children (Engel de Abreu, Conway, & Gathercole, 2010). The relation may reflect that both FLD-R and cWM measures involved concurrent storage and processing. Some researchers suggest that the storage of items that have been "bound" together to form some kind of association is responsible for the relation (Wilhelm, Hildebrandt, & Oberauer, 2013). Others argue it is CATT that is primarily responsible for the relationship (Shipstead, Harrison, & Engle, 2016).

With respect to comprehension, Andrews et al. (2017) have suggested that FLD-R captures the processing aspect of canonical sentence comprehension and perhaps both the processing and storage aspects of noncanonical sentence comprehension. Given our findings, it appears that FLD-R requires some degree of cWM, and the greater the relationship is between FLD-R and cWM, the better the comprehension of canonical and noncanonical sentences is. FLD-R was related to both the processing and storage aspects of cWM, and together, those abilities facilitated sentence comprehension. Finally, that our sentences lacked plausibility likely increased the importance of pattern recognition along with storage and processing working together to figure out and remember which thing mentioned in the sentence represented the agent (see next paragraph).

The cWM-mediated role of LTM-LK. cWM was also the conduit through which LTM-LK indirectly influenced comprehension. There is evidence that LTM and cWM are related but distinct constructs (Unsworth, 2010, 2016). We would again contend that cWM enabled the children to store the different linguistic chunks they had already created while the language system generated new chunks from new input. Moreover, in keeping with the spirit of Cowan's WM model (Cowan, 1999; Cowan et al., 2014), it seems reasonable to assume that the most recently generated chunk (e.g., verb phrase of main clause) was held in central storage (focal attention) whereas those chunks created from earlier processing occupied peripheral storage. A final sentence representation was then created by the language system as it combined the chunks from central storage and peripheral storage into a single, coherent representation.

In the developmental literature, it has been suggested that the poorer comprehension of noncanonical sentences compared with canonical sentences is attributable to storage differences between the two structures, with noncanonicals (not canonicals) involving the storage of two unintegrated NPs until processing the critical verb (Felser, Marinis, & Clahsen, 2003; Finney et al., 2014; Montgomery et al., 2008;

Weighall & Altmann, 2011). Our findings, however, lead us to a more nuanced view of the role of cWM. We would argue that the number of linguistic chunks does not underlie this comprehension difference. Rather, the critical factors appear to be the children's ability to chunk the input into different structurally relevant units (e.g., NVN, NNV) by activating in LTM those units with similar distributional characteristics as the input. All of our sentences contained just one or two chunks that were critical to comprehension. The SVOs [*The square had changed the bed* (*under the very new dry key*)] and passives [*The watch was bumped by the wheel* (*near the very bright clock*)] included just a single main clause. The subject relatives [*The watch [that had hugged the truck] (behind the kite) was bright*] and object relatives [*The chair [that the bread had splashed] (under the square) was new*] contained one main clause and one dependent clause. Even with the presence of a prepositional phrase near the end of the sentence, this chunk was not crucial to comprehension. The children apparently recognized this fact because when they miscomprehended a sentence, as noted above, they were more likely to choose the object noun as the agent, not the noun in the prepositional phrase. Thus, it appears that the chunking function growing out of the relationship between cWM and LTM-LK plays a crucial role in sentence comprehension, not the number of chunks per se.

The interpretation that cWM storage was not responsible for the poorer comprehension of noncanonical sentences aligns well with recent memory-based frameworks advanced in the adult sentence comprehension literature in which it is argued that comprehenders typically have plenty of storage to retain the necessary linguistic chunks needed to generate a complex sentence representation (Lewis, Vasishth, & Van Dyke, 2006; Van Dyke & McElree, 2006, 2011). Instead, a major determinant of successful comprehension is the ability to resolve any similarity-based retrieval interference, that is, the ability to selectively reactivate NP1 over NP2 upon encountering the critical verb (*The defendant that the lawyer talked with was charged with a felony*). If the cues available at retrieval (critical verb) are not sufficiently syntactically and/or semantically distinctive to facilitate the reactivation of NP1 over NP2, then retrieval interference arises. Under these conditions, NP2 may be retrieved leading to improper development of a filler-gap dependency and semantic role assignments to the NPs. However, when sufficient cues are available, retrieval interference is reduced or even eliminated (Van Dyke & McElree, 2011). Because our noncanonical sentences were designed to be implausible, they, by default, led to significant retrieval interference. Although the canonical sentences could be processed linearly, we might argue that their implausibility led to a global interference effect with respect to making proper semantic role assignments.

The nonrole of the cWM-mediated influence of CATT. We envisioned that children's CATT abilities may be important to the comprehension of our implausible sentences. Sustained attention may facilitate children's ability to maintain attention over the course of a sentence by focusing their mental energies on incoming lexical items. The ability to switch attention efficiently may facilitate the ability to toggle between storing linguistic chunks that are being created in the language system while continuing to process new input. That CATT had no significant indirect influence (via cWM) on comprehension implies that the children comprehended the canononical and noncanonical sentences with some degree of automaticity. Even in the absence of plausibility, TD children's sentence comprehension apparently occurs more or less automatically due to chunking efficiency in LTM, with little to no involvement of CATT. This interpretation aligns well with both the good-enough sentence comprehension model that emphasizes fast and frugal processing of input and the chunk-and-pass model that stresses the immediate chunking of input and the activation of relevant chunks from memory. An interpretation of relative automaticity is in keeping with arguments we have made elsewhere (Ahmad Rusli & Montgomery, 2017; Montgomery & Evans, 2009; Montgomery et al., 2009).

DLD Group

The cWM mediation model fit was good for both the TD and DLD groups, suggesting that the same overall relationship between cognitive processing and sentence comprehension held for both groups. But the magnitude of the indirect influences of FLD-R, LTM-LK, and CATT varied between the groups, leading to different mechanisms underlying sentence comprehension in children with DLD and their TD mates.

The cWM-mediated role of LTM-LK. The significant indirect influence of LTM-LK on the comprehension of children with DLD was structure dependent; its influence was restricted to canonical structures, implying that these children possessed structurally based multiword representations (NVN sequence) for these structures. The contribution of LTM-LK to canonical sentence comprehension, interestingly, was highly similar for both the DLD and TD groups. Such findings suggest that the children with DLD had the ability to chunk canonical input into relevant units such as phrases or even larger NVN units to construct a canonical representation. This interpretation is in keeping with both the chunk-and-pass and good-enough perspectives. We would also submit that the children with DLD had enough cWM capacity to support canonical sentence comprehension given their mean span score on the auditory cWM measure was 2.6 and the canonical structures comprised just one clause (SVOs) or two clauses (subject relatives). The children with DLD were thus likely able to use basic chunking abilities to aid in the creation of canonical sentence representations using just structural cues, but not as well as TD children. One alternative explanation for the cWM-mediated influence of LTM-LK on the canonical sentence comprehension by the DLD group is that these children processed the sentences not through chunking but on a word-by-word basis. But we reject this possibility because such an approach would have led to a memory overload, as the children would have had to store six or seven words before determining the agent with some confidence.

For the noncanonical structures, LTM-LK had no appreciable indirect influence on comprehension. In the case of passives, even though they conform to an NVN word order, they include key cues such as an inflected verb and a *by*-phrase marking them as noncanonical. Object relatives, on the other hand, involve a different word order all together—NNV. That LTM-LK had no effect on comprehension suggests that the learning of and memory for noncanonical multiword sequences is a severe weakness in children with DLD. This interpretation is consistent with claims that individuals with DLD have difficulty in learning the statistical properties of auditory input (Evans, Saffran, & Robe-Torres, 2009; Grunow, Spaulding, Gómez, & Plante, 2006; Hsu & Bishop, 2010), including word order (Plante, Gomez, & Gerken, 2002). Such findings also suggest that a chunk-and-pass processing approach was ineffective for the children, leading them to apply a simple NP1 as agent strategy (Evans, 2002; Evans & MacWhinney, 1999) to create a default and incorrect NVN representation (consistent with a good-enough approach; Ferreira et al., 2002; Karimi & Ferreira, 2016). Evidence for this interpretation comes from the finding that the children incorrectly selected the object as the agent of the sentence and not the noun in the prepositional phrase (Montgomery et al., 2017).

Relative to same-age peers, children with DLD may be much more dependent on the co-occurrence of semantic–pragmatic cues to learn multiword chunks corresponding to noncanonical structures to promote their extraction and abstraction of less frequently occurring word order patterns. The children may even need the strongest of semantic–pragmatic cues to promote the learning of complex, noncanonical structures. The finding that children with DLD and age peers show comparable comprehension of *be* passives when strong semantic–pragmatic cues are available (*The milk was spilled by the girl*; van der Lely, 1998) would seem to illustrate this possibility. If it can be shown that noncanonical sentence comprehension by children with DLD is comparable to or approaches the level of TD children, such findings would shed new light on the grammar learning abilities of children with DLD and would seem to have implications for the procedural memory deficit hypothesis of DLD (Lum, Conti-Ramsden, Page, & Ullman, 2012; Ullman & Pierpont, 2005), which posits that the grammar learning problems of these children is related to a deficit in learning sequential patterns.

The cWM-mediated role of CATT. CATT operating through cWM also influenced the comprehension of the children with DLD. We hypothesized that the children's CATT abilities should be important to their comprehension, especially given the implausibility of the sentences. Sustained attention should be important because it would allow the children to attend to incoming lexical material over the course of the input. We also reasoned that attention switching should be important because it would permit the children to toggle their focal attention between storing chunked units (e.g., NP, verb phrase) in memory from earlier processing while allowing the language system to process/chunk new, incoming information (e.g., NPs). That CATT played a significant indirect role in simple sentence comprehension of children with DLD is consistent with these assumptions. Such findings imply that canonical comprehension in these children is a cognitively effortful and inefficient process. This interpretation is consistent with findings of previous DLD studies investigating these children's canonical sentence comprehension (Lum, Youssef, & Clark, 2017; Montgomery, 2000; Montgomery & Evans, 2009; Montgomery et al., 2009).

CATT also indirectly influenced the comprehension of noncanonical structures by the DLD group. That is, for the children in the DLD group, better sustained attention and attention switching was related to better cWM, which led to stronger noncanonical sentence comprehension. We assume that sustained attention played the same role in these sentences as in the canonical sentences. Given the nonrole of LTM-LK, attention switching likely involved a simple toggling of attention between storage (i.e., storing as many individual words as possible encountered earlier in the sentence) and attempts to process new, incoming language material. In other words, faced with weak noncanonical word order knowledge and limited linguistic chunking abilities for these structures, the children may have relied on a simple word-by-word processing approach that played out in the NP1 as agent strategy (Evans, 2002; Evans & MacWhinney, 1999). In the absence of viable multiword representations in LTM and poor multiword chunking abilities for these structures, a simple yet ineffectual and cognitively effortful word-by-word processing strategy may have been all that was available to the children. An interpretation of cognitive effort underlying more complex sentence comprehension is also consistent with previous claims suggested in the DLD literature (Montgomery, 2008; Montgomery & Evans, 2009; Montgomery et al., 2009).

The nonrole of the cWM-mediated influence of FLD-R. FLD-R had no significant indirect influence on the comprehension of either canonical or noncanonical structures in the children with DLD. Recall that FLD-R and syntactic processing may be similar because each involves the ability to recognize and interpret patterns in the input (Andrews et al., 2017). The nonrole of FLD-R in the children with DLD indicates that general pattern recognition and analytic reasoning abilities play no appreciable role in supporting these children's comprehension. Such findings suggest that, in the absence of the availability of general pattern recognition and reasoning abilities, sentence comprehension in children with DLD is made more difficult.

How to Characterize Sentence Comprehension of the Children With DLD: Categorically or Dimensionally?

There is some debate as to whether DLD is better conceptualized as a discrete, categorical disorder or a dimensional disorder (e.g., Bishop, 2017). A categorical approach to defining a disorder assumes that the observed symptoms of the disorder are either present or absent based on performance on one or more tests, with presence or absence of disorder defined according to an arbitrary cutoff score (i.e., standard score, scaled score) on one or more tests. However, such an approach does not take into consideration different symptom profiles, symptom severity, or underlying factors that may contribute to the disorder, which may lead to failure to characterize the heterogeneity and the complexity of symptoms and factors associated with the disorder. A dimensional approach assumes that ability falls along a continuum, with the language abilities of children with DLD falling at the low end of the normal distribution. One major problem with this approach is that, across different language measures, children with DLD may show a range of performance that is below or above the low end of the distribution, leading to problems correctly identifying a child as DLD (Plante & Vance, 1994; Spaulding, Plante, & Farinella, 2006).

The question might be asked whether the syntactic comprehension (and more precisely the structural relationship between cognitive processing and syntactic comprehension) observed in our children with DLD aligns better with a categorical view of language ability or a dimensional view. On the one hand, the full group SEM results indicated that the structural relationship was similar for both the DLD and TD groups, suggesting that sentence comprehension in both groups was indirectly influenced by the same cognitive mechanisms. However, recall that the model fit statistics for the multigroup model that accounted for variation in the two groups were better than the fit statistics when all the children were combined. The model in which cWM played a mediating role was a better characterization of the cognitive contributions to sentence comprehension when we controlled for the variance that was attributable to each group separately. In addition, there were important and theoretically meaningful differences between the groups with respect to the magnitude of the indirect influences of FLD-R, CATT, and LTM-LK. Such variations suggest that there were critical and nuanced differences between the groups in the role played by these cognitive mechanisms.

Summary

We offer the first theoretically integrated and empirically validated cognitive model of the syntactic sentence comprehension abilities of a large group of school-age children with DLD and a propensity-matched group of TD children. The GEM model offers substantively new and important insights into the sentence comprehension abilities of children with DLD. It is clear that LTM-LK, CATT, and FLD-R do not affect comprehension independently of each other. Rather these mechanisms appear to affect sentence comprehension by way of their relationships with cWM. For TD children, deeper and stronger LTM-LK (crystallized knowledge), together with better general pattern recognition abilities, leads to more efficient cWM through automatic chunking of the input, which, in turn, leads to serviceable canonical and noncanonical sentence comprehension.

For children with DLD, sentence comprehension is an effortful process for both canonical and noncanonical structures. General pattern recognition abilities appear to play no appreciable role in comprehension. For canonical structures, though, these children do appear to use their LTM-LK to chunk the input into relevant linguistic units, but they do so with less accuracy and greater cognitive effort than TD children. For noncanonical sentences, however, the children in the DLD group appear to comprehend the input on a word-by-word basis.

An important and novel strength of this study is that our groups were propensity-matched for age, gender, mother's education, and family income. Adding children's propensity scores as a covariate in our models enabled us to calculate standardized *YX* values for the two groups that were not biased by the effects that age, gender, and SES might have on the relationships between cognition and sentence comprehension. Incorporating propensity scoring into the study design extends the degree of confidence that the comprehension differences between the groups and the different cognitive mechanisms underlying each group's comprehension are likely to generalize to other 7- to 11-year-old children, regardless of gender or SES background.

How Do Our Findings Align With Other Models of cWM?

We adopted Cowan's WM model because of its integrated processes view of attention, activation of information from LTM in a capacity-limited store, and the explicit chunking function of LTM. But do our findings and interpretations align with other models of WM? We focus on storage here. The dual-storage model of Engle and colleagues (Engle, Tuholski, Laughlin, & Conway, 1999; Kane & Engle, 2002; Kane et al., 2004) would appear to be a reasonable fit with our data. Similar to Cowan's model, the dual-store model assumes cWM capacity to reflect the combination of items occupying primary memory (central storage) and secondary memory (peripheral storage). It seems reasonable to assume that primary memory would hold the most recently generated linguistic chunk whereas secondary memory holds those chunks created from earlier processing.

Oberauer and Hein (2012) have proposed a three embedded components model, which comprises an activated portion of LTM, broad focus of attention, and a narrow focus of attention. The broad focus of attention, also referred to as the region of direct access, includes those items that have been taken in from all the items that have just been activated in LTM. The broad focus of attention (roughly analogous to peripheral storage) is limited to about four items. The narrow focus of attention (roughly analogous to central storage) is limited to a single item. This model appears consistent with our interpretation in that the narrow focus of attention holds the linguistic chunk that has been most recently created and that the broad focus of attention holds chunks from earlier processing.

Baddeley's (2012) multicomponent model of WM, however, is not well represented by our findings, given the measures we employed. The relevant components of the Baddeley model to spoken sentence comprehension include the phonological loop, the episodic buffer, and executive attention (e.g., attention allocation, sustained attention, divided attention). As in Cowan's model, WM interfaces with LTM because those items that have been activated reside in LTM. The phonological loop temporarily stores speech input but is severely limited to a few items (e.g., words). The episodic buffer serves as an interface between the phonological loop and LTM. The buffer is responsible for storing information units that are beyond the loop's capacity—such as large, integrated chunks (e.g., clauses). It is the language system that creates these chunks by binding words into clauses and a final sentence representation. The buffer simply stores these representations. Had we included an episodic buffer construct and the buffer proved to have a direct or indirect influence on the children's comprehension, then the multicomponent model, too, would be a good fit.

Limitations of the Current Study

We conducted an exploratory study of the structural relationship between a set of latent cognitive predictors and sentence comprehension controlling for the factors contributing to our propensity score. We estimated cWM with a standardized word-level measure and an experimental measure. It is possible that different types of cWM measures such as listening span tasks with sentence-level stimuli might have led to different estimations of the extent of the direct relationships among FLD-R, CATT, LTM-LK, and cWM, as well as the extent to which they all related indirectly to sentence comprehension.

We built our GEM measurement model to be consistent with empirical and theoretical evidence of the cognitive mechanisms associated with sentence comprehension in children with and without DLD. It is possible that there are other mechanisms that we did not include (e.g., inhibition of interference) that could alter or negate the role of cWM as a mediator. We are aware that our claims relate only to the role of cWM when assessed by the memory measures we used and when modeled with the covariates that we included. We recognize, for example, that the measure we used to index LTM-LK, the Test of Narrative Language, was a rather broad measure encompassing both sentence-level and discourse-level (re)construction abilities. Had we used LTM-LK measures that were more proximal to sentence comprehension (e.g., lexical knowledge, sentence-level grammatical knowledge), our model(s) may have been different. For instance, we may have seen LTM-LK play a more reliable role in the sentence comprehension of the children with DLD.

Finally, in our explanation of the mediator role of cWM, we had to assume a temporal order between the mediator and sentence comprehension. A true causal test of our hypothesis that cWM mediates the influence of the other cognitive mechanisms on children's sentence comprehension requires a counterfactual approach (Vander Weele, 2015) to modeling the results of a randomized treatment study in which the model includes a measure of cWM prior to intervention along with measures of sentence comprehension after intervention.

Future Research Directions

Future research directions are rich. Researchers may wish to explore the learning of different multiword patterns as a function of the availability of different semantic–pragmatic cues and cue combinations in children with DLD. This approach would enable us to evaluate our claim that the learning of multiword units by these children is dependent on the presence of strong semantic–pragmatic cues. If it can be shown that the availability of strong semantic–pragmatic cues in noncanonical structures leads to reliably better comprehension, such findings would yield new and important theoretical and clinical implications about the grammar learning abilities of these children and the conditions that promote stable learning/retention of complex word order patterns. An important, related issue would be to determine whether such learning translates to more automatic multiword chunking of input, and if so, whether greater automaticity leads to a different relationship between cognitive processing and comprehension.

Our current findings describe the structural relationship between cognitive processing and sentence comprehension in school-age children with and without DLD. An obvious question is whether this relationship holds across a wider age range or whether it may change with age. Adolescents with DLD, for example, having accrued more language processing experience, may acquire stronger representations of both noncanonical and canonical multiword representations. If so, we might expect language LTM to play an even stronger role in adolescent comprehension than it does during the school-age years, that is, more automatic and accurate linguistic chunking of the input, for different structures containing a range of different linguistic cues. Finally, researchers can use SEM to investigate whether the sentence comprehension deficit of children with DLD is better characterized as a categorical or dimensional deficit.

If cWM is a true mediator of controlled (executive) attention and this mediated relationship supports sentence comprehension, it should be possible through neuroimaging techniques to identify those brain regions typically associated with verbal cWM (e.g., left prefrontal cortex, anterior cingulate cortex, temporal language area; Chein, Moore, & Conway, 2011; Osaka et al., 2003) and CATT related to cWM performance (dorsolateral prefrontal cortex, frontoparietal system; Peterson & Posner, 2012) and assess the timing of coactivation of these regions during language processing. As noted in the limitations section, another approach would be to model the effects of training LTM-LK, CATT, or both on children's sentence comprehension abilities using pretest measures of cWM as the mediator. This approach would be consistent with recommendations from some investigators (Bishop, 2009) that stress the need to consider the cognitive processing deficiencies of these children when treating their language problems. Neuroimaging studies could also be combined with experimental studies to look for changes in strength and timing of neural (co)activation following different kinds of training exposures. Finally, future studies may examine the relationship between cognitive processing and real-time sentence processing to broaden and deepen our understanding of DLD sentence comprehension.

Acknowledgments

This study was funded by Grant R01 DC010883 from the National Institute on Deafness and Other Communicative Disorders, awarded to Julia L. Evans, Ronald B. Gillam, and James W. Montgomery. Special thanks to Alexander Sergeev, who conducted the propensity matching. We express our gratitude to all the children and their parents who participated in this project. We also thank Beula Magimairaj, Naveen Nagaraj, Misha Finney, Yazmine Ahmad Rusli, Jenny Boyden, Andrea Fung, Katie Squires, Kelly Rogers, Llely Duarte, Allison Hancock, and Farzaneh Vahabi for their assistance during various phases of this study.

References

Abbot-Smith, K., & Tomasello, M. (2006). Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *The Linguistic Review, 23,* 275–290. https://doi.org/10.1515/TLR.2006.011

Ahmad Rusli, Y., & Montgomery, J. W. (2017). Children's comprehension of object relative sentences: It's extant language knowledge that matters, not domain-general working memory. *Journal of Speech, Language, and Hearing Research, 60,* 2865–2878. https://doi.org/10.1044/2017_JSLHR-L-16-0422

Alonzo, C., Yeomans-Maldonado, G., Murphy, K., & Bevens, B. (2016). Predicting second grade listening comprehension using prekindergarten measures. *Topics in Language Disorders, 36,* 312–333. https://doi.org/10.1097/TLD.0000000000000102

American National Standards Institute. (1997). *Specifications of audiometers* (ANSI/ANS 8.3-1997, R2003). New York, NY.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Andrews, G., Ogden, J. E., & Halford, G. S. (2017). Resolving conflicts between syntax and plausibility in sentence comprehension. *Advances in Cognitive Psychology, 13,* 11–27. https://doi.org/10.5709/acp-0203-8

Archibald, L. M. D., & Gathercole, S. E. (2007). The complexities of complex memory span: Storage and processing deficits in specific language impairment. *Journal of Memory and Language, 57,* 177–194. https://doi.org/10.1016/j.jml.2006.11.004

Arnon, I., & Christiansen, M. H. (2017). The role of multiword building blocks in explaining L1–L2 differences. *Topics in Cognitive Science, 9,* 621–636. https://doi.org/10.1111/tops.12271

Arnon, I., & Clark, E. V. (2011). Why brush your teeth is better than teeth—Children's word production is facilitated in familiar sentence-frames. *Language Learning and Development, 7,* 107–129. https://doi.org/10.1080/15475441.2010.505489

Arnon, I., McCauley, S. M., & Christiansen, M. H. (2017). Digging up the building blocks of language: Age-of-acquisition effects for multiword phrases. *Journal of Memory and Language, 92,* 265–280. https://doi.org/10.1016/j.jml.2016.07.004

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language, 62,* 67–82. https://doi.org/10.1016/j.jml.2009.09.005

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63,* 1–29. https://doi.org/10.1146/annurev-psych-120710-100422

Bannard, C., & Lieven, E. (2012). Formulaic language in L1 acquisition. *Annual Review of Applied Linguistics, 32,* 3–16. https://doi.org/10.1017/S0267190512000062

Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning. *Psychological Science, 19,* 241–248. https://doi.org/10.1111/j.1467-9280.2008.02075.x

Barrett, P. (2007). Structural equation modelling: Judging model fit. *Personality and Individual Differences, 42,* 815–824. https://doi.org/10.1016/j.paid.2006.09.018

Barrouillet, P., Gavens, N., Vergauwe, E., Gaillard, V., & Camos, V. (2009). Working memory span development: A time-based resource-sharing model account. *Developmental Psychology, 45,* 277–290. https://doi.org/10.1016/j.paid.2006.09.018

Bates, E., & MacWhinney, B. (1987). Competition variation and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 157–193). Hillsdale, NJ: Erlbaum.

Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney & E. Bates (Eds.), *The cross-linguistic study of sentence processing* (pp. 3–73). Cambridge, United Kingdom: Cambridge University Press.

Bedore, L. M., Peña, E. D., Summers, C. L., Boerger, K. M., Resendiz, M. D., Greene, K., . . . Gillam, R. B. (2012). The measure matters: Language dominance profiles across measures in Spanish–English bilingual children. *Bilingualism: Language and Cognition, 15,* 616–629. https://doi.org/10.1017/S1366728912000090

Bencini, G. M. L., & Valian, V. V. (2008). Abstract sentence representations in 3-year-olds: Evidence from language production and comprehension. *Journal of Memory and Language, 59,* 97–113. https://doi.org/10.1016/j.jml.2007.12.007

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107,* 238–246. https://doi.org/10.1037/0033-2909.107.2.238

Bentler, P. M., & Bonnet, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88,* 588–606. https://doi.org/10.1037/0033-2909.88.3.588

Bishop, D. V. M. (2009). Genes, cognition, and communication: Insights from neurodevelopmental disorders. *The Year in Cognitive Neuroscience, 1156,* 1–18. https://doi.org/10.1111/j.1749-6632.2009.04419.x

Bishop, D. V. M. (2017). Why is it so hard to reach agreement on terminology? The case of developmental language disorder (DLD). *International Journal of Language and Communication Disorders, 52,* 671–680. https://doi.org/10.1111/1460-6984.12335

Boyle, W., Lindell, A. K., & Kidd, E. (2013). Investigating the role of verbal working memory in young children's sentence comprehension. *Language Learning, 63,* 211–242. https://doi.org/10.1111/lang.12003

Browne, M., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. Bollen & J. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Burgess, G. C., Gray, J. R., Conway, A. R., & Braver, T. S. (2011). Neural mechanisms of interference control underlie the relationship between fluid intelligence and working memory span. *Journal of Experimental Psychology: General, 140,* 674–692. https://doi.org/10.1037/a0024695

Chater, N., McCauley, S. M., & Christiansen, M. H. (2016). Language as skill: Intertwining comprehension and production. *Journal of Memory and Language, 89,* 244–254. https://doi.org/10.1016/j.jml.2015.11.004

Chein, J. M., Moore, A. B., & Conway, A. R. (2011). Domain-general mechanisms of complex working memory span. *NeuroImage, 54,* 550–559. https://doi.org/10.1016/j.neuroimage.2010.07.067

Christiansen, M. H., & Arnon, I. (2017). More than words: The role of multiword sequences in language learning and use. *Topics in Cognitive Science, 9,* 542–551. https://doi.org/10.1111/tops.12274

Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences, 39,* e62. https://doi.org/10.1017/S0140525X1500031X

Christiansen, M. H., & MacDonald, M. C. (1999). Fractionated working memory: Even in pebbles, it's still a soup stone. *Behavioral and Brain Sciences, 22,* 97–98. https://doi.org/10.1017/S0140525X99251783

Conti-Ramsden, G., Ullman, M. T., & Lum, J. A. G. (2015). The relation between receptive grammar and procedural, declarative, and working memory in specific language impairment. *Frontiers in Psychology, 6,* 1090. https://doi.org/10.3389/fpsyg.2015.01090

Cornish, H., Dale, R., Kirby, S., & Christiansen, M. H. (2017). Sequence memory constraints give rise to language-like structure through iterated learning. *PLoS One, 12,* e0168532. https://doi.org/10.1371/journal.pone.0168532

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Model of working memory: Mechanism of active maintenance and executive control* (pp. 62–101). New York, NY: Cambridge University Press.

Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology, 51,* 42–100. https://doi.org/10.1016/j.cogpsych.2004.12.001

Cowan, N., Fristoe, N. M., Elliott, E. M., Brunner, R. P., & Saults, J. S. (2006). Scope of attention, control of attention, and intelligence in children and adults. *Memory and Cognition, 34,* 1754–1768. https://doi.org/10.3758/BF03195936

Cowan, N., Rouder, J. N., Blume, C. L., & Saults, J. S. (2012). Models of verbal working memory capacity: What does it take to make them work? *Psychological Review, 119,* 480–499. https://doi.org/10.1037/a0027791

Cowan, N., Saults, J. S., & Blume, C. L. (2014). Central and peripheral components of working memory storage. *Journal of Experimental Psychology: General, 143,* 1806–1836. https://doi.org/10.1037/a0036814

D'Agostino, R. B., Jr. (1998). Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine, 17,* 2265–2281. https://doi.org/10.1002/(SICI)1097-0258(19981015)17:19<2265::AID-SIM918>3.0.CO;2-B

Dick, F., Wulfeck, B., Krupa-Kwiatkowski, M., & Bates, E. (2004). The development of complex sentence interpretation in typically developing children compared with children with specific language impairment or early unilateral focal lesions. *Developmental Science, 7,* 360–377. https://doi.org/10.1111/j.1467-7687.2004.00353.x

Ellis Weismer, S., Evans, J., & Hesketh, L. J. (1999). An examination of verbal working memory capacity in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 42,* 1249–1260. https://doi.org/10.1044/jslhr.4205.1249

Engel de Abreu, P. M. J., Conway, A. R. A., & Gathercole, S. E. (2010). Working memory and fluid intelligence in young children. *Intelligence, 38,* 552–561. https://doi.org/10.1016/j.intell.2010.07.003

Engelhardt, P. E., Nigg, J. T., & Ferreira, F. (2017). Executive function and intelligence in the resolution of temporary syntactic ambiguity: An individual differences investigation. *The Quarterly Journal of Experimental Psychology, 70,* 1263–1281. https://doi.org/10.1080/17470218.2016.1178785

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General, 128,* 309–331. https://doi.org/10.1037/0096-3445.128.3.309

Evans, J. L. (2002). Variability in comprehension strategy use in children with specific language impairments: A dynamical systems account. *International Journal of Language and Communication Disorders, 37,* 95–116. https://doi.org/10.1080/13682820110116767

Evans, J. L., & MacWhinney, B. (1999). Sentence processing strategies in children with expressive and expressive–receptive specific language impairments. *International Journal of Language and Communication Disorders, 34,* 117–134. https://doi.org/10.1080/136828299247469

Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 52,* 321–335. https://doi.org/10.1044/1092-4388(2009/07-0189)

Felser, C., Marinis, T., & Clahsen, H. (2003). Children's processing of ambiguous sentences: A study of relative clause attachment. *Language Acquisition, 11,* 127–163. https://doi.org/10.1207/s15327817la1103_1

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11,* 11–15. https://doi.org/10.1111/1467-8721.00158

Ferreira, F., & Patson, N. D. (2007). The good-enough approach to language comprehension. *Language and Linguistics Compass, 1,* 71–83. https://doi.org/10.1111/j.1749-818x.2007.00007.x

Finney, M. C., Montgomery, J. W., Gillam, R. B., & Evans, J. L. (2014). Role of working memory storage and attention focus switching in children's comprehension of spoken object relative sentences. *Child Development Research, 54,* 1–11. https://doi.org/10.1155/2014/450734

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. https://doi.org/10.1017/S0305000904006269

Friedmann, N., & Novogrodsky, R. (2007). Is the movement deficit in syntactic SLI related to traces or to thematic role transfer? *Brain and Language, 101,* 50–63. https://doi.org/10.1016/j.bandl.2006.09.006

Fukuda, K., Vogel, E., Mayr, U., & Awh, E. (2010). Quantity, not quality: The relationship between fluid intelligence and working memory capacity. *Psychonomic Bulletin & Review, 17,* 673–679. https://doi.org/10.3758/17.5.673

García-Madruga, J. A., Vila, J. O., Gómez-Veiga, I., Duque, G., & Elosúa, M. R. (2014). Executive processes, reading comprehension and academic achievement in 3rd grade primary students. *Learning and Individual Differences, 35,* 41–48. https://doi.org/10.1016/j.lindif.2014.07.013

Gilchrist, A. L., Cowan, N., & Naveh-Benjamin, M. (2009). Investigating the childhood development of working memory using sentences: New evidence for the growth of chunk capacity. *Journal of Experimental Child Psychology, 104,* 252–265. https://doi.org/10.1016/j.jecp.2009.05.006

Gillam, R. B., & Johnston, J. R. (1992). Spoken and written language relationships in language/learning-impaired and normally achieving school-age children. *Journal of Speech and Hearing Research, 35*(6), 1303–1315.

Gillam, R. B., & Pearson, N. (2004). *Test of Narrative Language*. Austin, TX: Pro-Ed.

Grunow, H., Spaulding, T. J., Gómez, R. L., & Plante, E. (2006). The effects of variation on learning word order rules by adults with and without language-based learning disabilities. *Journal of Communication Disorders, 39,* 158–170. https://doi.org/10.1016/j.jcomdis.2005.11.004

Haavisto, M.-L., & Lehto, J. E. (2005). Fluid/spatial and crystallized intelligence in relation to domain-specific working memory: A latent-variable approach. *Learning and Individual Differences, 15,* 1–21. https://doi.org/10.1016/j.lindif.2004.04.002

Hsu, H. J., & Bishop, D. V. M. (2010). Grammatical difficulties in children with specific language impairment: Is learning deficient? *Human Development, 53,* 264–277. https://doi.org/10.1159/000321289

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria vs. new alternatives. *Structural Equation Modeling, 6,* 56–83. https://doi.org/10.1080/10705519909540118

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99,* 122–149. https://doi.org/10.1037/0033-295X.99.1.122

Just, M. A., Carpenter, P. A., & Keller, T. A. (1996). The capacity theory of comprehension: New frontiers of evidence and arguments. *Psychological Review, 103,* 773–780. https://doi.org/10.1037/0033-295X.103.4.773

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review, 9,* 637–671. https://doi.org/10.3758/BF03196323

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133,* 189–217. https://doi.org/10.1037/0096-3445.133.2.189

Karimi, H., & Ferreira, F. (2016). Good-enough linguistic representations and online cognitive equilibrium in language processing. *The Quarterly Journal of Experimental Psychology, 69,* 1013–1040. https://doi.org/10.1080/17470218.2015.1053951

Kim, A., & Sikos, L. (2011). Conflict and surrender during sentence processing: An ERP study of syntax-semantics interaction. *Brain and Language, 118,* 15–22. https://doi.org/10.1016/j.bandl.2011.03.002

Kintsch, W., Patel, V. L., & Ericsson, K. A. (1999). The role of long-term working memory in text comprehension. *Psychologia: An International Journal of Psychology in the Orient, 42*(4), 186–198.

Kline, R. (2016). *Principles and practices of structural equation modeling* (4th ed.). New York, NY: Guilford Press.

Kuperberg, G. R., Kreher, D. A., Sitnikova, T., Caplan, D. N., & Holcomb, P. J. (2007). The role of animacy and thematic relationships in processing active English sentences: Evidence from event-related potentials. *Brain and Language, 100,* 223–237. https://doi.org/10.1016/j.bandl.2005.12.006

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4,* 463–470. https://doi.org/10.1016/S1364-6613(00)01560-6

Leonard, L. B. (2014). *Children with specific language impairment* (2nd ed.). Cambridge, MA: MIT Press.

Leonard, L. B., Deevy, P., Fey, M. E., & Bredin-Oja, S. L. (2013). Sentence comprehension in specific language impairment: A task designed to distinguish between cognitive capacity and syntactic complexity. *Journal of Speech, Language, and Hearing Research, 56,* 577–589. https://doi.org/10.1044/1092-4388(2012/11-0254)

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences, 10,* 447–454. https://doi.org/10.1016/j.tics.2006.08.007

Lieven, E., Salomo, D., & Tomasello, M. (2009). Two-year-old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics, 20,* 481–507. https://doi.org/10.1515/COGL.2009.022

Lum, J., Conti-Ramsden, G., & Lindell, A. (2007). The attentional blink reveals sluggish attention shifting in adolescents with specific language impairment. *Brain Cognition, 63,* 287–295. https://doi.org/10.1016/j.bandc.2006.09.010

Lum, J., Conti-Ramsden, G., Page, D., & Ullman, M. (2012). Working, declarative and procedural memory in specific language impairment. *Cortex, 48,* 1138–1154. https://doi.org/10.1016/j.cortex.2011.06.001

Lum, J., Youssef, G., & Clark, G. (2017). Using pupillometry to investigate sentence comprehension in children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research, 60,* 1648–1660. https://doi.org/10.1044/2017_JSLHR-L-16-0158

Marton, K., Campanelli, L., Eichorn, N., Scheuer, J., & Yoon, J. (2014). Information processing and proactive interference in children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research, 57,* 106–119. https://doi.org/10.1044/1092-4388(2013/12-0306)

McCauley, S. M., & Christiansen, M. H. (2013). Toward a unified account of comprehension and production in language development. *Behavioral and Brain Sciences, 36,* 38–39. https://doi.org/10.1017/S0140525X12002658

McCauley, S. M., & Christiansen, M. H. (2014). Acquiring formulaic language: A computational model. *The Mental Lexicon, 9,* 419–436. https://doi.org/10.1075/ml.9.3.03mcc

McCauley, S. M., & Christiansen, M. H. (2015). Individual differences in chunking ability predict on-line sentence processing. In D. C. Noelle & R. Dale (Eds.), *Proceedings from the 37th Annual Conference of the Cognitive Science Society* (pp. 1553–1558). Austin, TX: Cognitive Science Society. https://mindmodeling.org/cogsci2015/papers/0271/paper0271.pdf

McCauley, S. M., & Christiansen, M. H. (2017). Computational investigations of multiword chunks in language learning. *Topics in Cognitive Science, 9,* 637–652. https://doi.org/10.1111/tops.12258

McCauley, S. M., Isbilen, E., & Christiansen, M. H. (2017). Chunking ability shapes sentence processing at multiple levels of abstraction. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 2681–2686). Austin, TX: Cognitive Science Society. https://www.researchgate.net/publication/316926923

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum.

McElree, B. (2006). Accessing recent events. *Psychology of Learning and Motivation, 46,* 155–200. https://doi.org/10.1016/S0079-7421(06)46005-9

Montgomery, J. W. (2000). Verbal working memory and sentence comprehension in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 43,* 293–308. https://doi.org/10.1044/jslhr.4302.293

Montgomery, J. W. (2008). Role of auditory attention in the real-time processing of simple grammar by children with specific language impairment: A preliminary investigation. *International Journal of Language and Communication Disorders, 43,* 499–527. https://doi.org/10.1080/13682820701736638

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. https://doi.org/10.1044/1092-4388(2008/07-0116)

Montgomery, J. W., Evans, J. L., & Gillam, R. B. (2009). Relation of auditory attention and complex sentence comprehension in children with specific language impairment: A preliminary study. *Applied Psycholinguistics, 30,* 123–151. https://doi.org/10.1017/S0142716408090061

Montgomery, J. W., Evans, J. L., Gillam, R. B., Sergeev, A. V., & Finney, M. C. (2016). "Whatdunit?" Developmental changes in children's syntactically based sentence comprehension abilities and sensitivity to word order. *Applied Psycholinguistics, 37,* 1281–1309. https://doi.org/10.1017/S0142716415000570

Montgomery, J. W., Gillam, R. B., Evans, J. L., & Sergeev, A. V. (2017). Whatdunit? Sentence comprehension abilities of children with SLI: Sensitivity to word order in canonical and non-canonical sentences. *Journal of Speech, Language, and Hearing Research, 60,* 2603–2618. https://doi.org/10.1044/2017_JSLHR-L-17-0025

Montgomery, J. W., Magimairaj, B. M., & O'Malley, M. H. (2008). The role of working memory in typically developing children's complex sentence comprehension. *Journal of Psycholinguistic Research, 37,* 331–354. https://doi.org/10.1007/s10936-008-9077-z

Motallebzadeh, K., & Yazdi, M. T. (2016). The relationship between EFL learners' reading comprehension ability and their fluid intelligence, crystallized intelligence, and processing speed. *Cogent Education, 3,* 1–8. https://doi.org/10.1080/2331186X.2016.1228733

Muthén, B. O., Muthén, L. K., & Asparouhov, T. (2016). *Regression and mediation analysis using Mplus*. Los Angeles, CA: Muthén & Muthén.

Newcomer, P., & Hammill, D. (2008). *Test of Language Development–Primary: Fourth Edition*. Austin, TX: Pro-Ed.

Norbury, C. F., Bishop, D. V. M., & Briscoe, J. (2002). Does impaired grammatical comprehension provide evidence of an innate grammar module? *Applied Psycholinguistics, 23,* 247–268. https://doi.org/10.1017/S0142716402002059

Oberauer, K., & Hein, L. (2012). Attention to information in working memory. *Current Directions in Psychological Science, 21,* 164–169. https://doi.org/10.1177/0963721412444727

Osaka, M., Osaka, N., Kondo, H., Morishita, M., Fukuyama, H., Aso, T., & Shibasaki, H. (2003). Neural basis of individual differences in working memory: An fMRI study. *NeuroImage, 18,* 789–797. https://doi.org/10.1016/S1053-8119(02)00032-0

Peterson, S. E., & Posner, M. I. (2012). The attention system of the human brain: 20 years after. *Annual Review of Neuroscience, 35,* 73–89. https://doi.org/10.1146/annurev-neuro-062111-150525

Plante, E., Gomez, R., & Gerken, L. (2002). Sensitivity to word order cues by normal and language/learning disabled adults. *Journal of Communication Disorders, 35,* 453–462. https://doi.org/10.1016/S0021-9924(02)00094-1

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

Reali, F., & Christiansen, M. H. (2007). Word-chunk frequencies affect the processing of pronominal object-relative clauses. *Quarterly Journal of Experimental Psychology, 60,* 161–170. https://doi.org/10.1080/17470210600971469

Robertson, E. K., & Joanisse, M. F. (2010). Spoken sentence comprehension in children with dyslexia and language impairment: The roles of syntax and working memory. *Applied Psycholinguistics, 31,* 141–165. https://doi.org/10.1017/S0142716409990208

Roid, G., & Miller, L. (1997). *Leiter International Performance Scale–Revised*. Wood Dale, IL: Stoelting.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70,* 41–55. https://doi.org/10.1093/biomet/70.1.41

Schneider, W., Eschman, A., & Zuccolott, A. (2002). *E-Prime user's guide*. Pittsburgh, PA: Psychology Software Tools.

Semel, E., Wiig, E., & Secord, W. (2003). *Clinical Evaluation of Language Fundamentals–Fourth Edition*. San Antonio, TX: The Psychological Corporation.

Shavers, V. L. (2007). Measurement of socioeconomic status in health disparities research. *Journal of the National Medical Association, 99,* 1013–1023.

Shipstead, Z., Harrison, T. L., & Engle, R. W. (2016). Working memory capacity and fluid intelligence: Maintenance and disengagement. *Perspectives on Psychological Science, 11,* 771–799. https://doi.org/10.1177/1745691616650647

Spaulding, T. J., Plante, E., & Farinella, K. A. (2006). Eligibility criteria for language impairment: Is the low end of normal always appropriate? *Language, Speech, and Hearing Services in Schools, 37,* 61–72. https://doi.org/10.1044/0161-1461(2006/007)

Steiger, J. H. (1990). Structural model evaluation and modification. *Multivariate Behavioral Research, 25,* 214–221. https://doi.org/10.1207/s15327906mbr2502_4

Tager-Flusberg, H., & Cooper, J. (1999). Present and future possibilities for defining a phenotype for specific language impairment. *Journal of Speech, Language, and Hearing Research, 42,* 1275–1278. https://doi.org/10.1044/jslhr.4205.1275

Theakston, A., & Lieven, E. (2017). Multiunit sequences in first language acquisition. *Topics in Cognitive Science, 9,* 588–603. https://doi.org/10.1111/tops.12268

Tomblin, J. B., Records, N. L., & Zhang, X. (1996). A system for the diagnosis of specific language impairment in kindergarten children. *Journal of Speech and Hearing Research, 39,* 1284–1294. https://doi.org/10.1044/jshr.3906.1284

Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning, 61,* 569–613. https://doi.org/10.1111/j.1467-9922.2010.00622.x

Ullman, M. T. (2001). A neurocognitive perspective on language: The declarative/procedural model. *Nature Reviews Neuroscience, 2,* 717–726. https://doi.org/10.1038/35094573

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition, 92,* 231–270. https://doi.org/10.1016/j.cognition.2003.10.008

Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex, 41,* 399–433. https://doi.org/10.1016/S0010-9452(08)70276-4

Unsworth, N. (2010). On the division of working memory and long-term memory and their relation to intelligence: A latent variable approach. *Acta Psychologica, 134,* 16–28. https://doi.org/10.1016/j.actpsy.2009.11.010

Unsworth, N. (2016). Working memory capacity and recall from long-term memory: Examining the influences of encoding strategies, study time allocation, search efficiency, and monitoring abilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42,* 50–61. https://doi.org/10.1037/xlm0000148

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114,* 104–132. https://doi.org/10.1037/0033-295X.114.1.104

Unsworth, N., & Robison, M. K. (2015). Individual differences in the allocation of attention to items in working memory: Evidence from pupillometry. *Psychonomic Bulletin and Review, 22,* 757–765. https://doi.org/10.3758/s13423-014-0747-6

Unsworth, N., & Robison, M. K. (2016). The influence of lapses of attention on working memory capacity. *Memory and Cognition, 44,* 188–196. https://doi.org/10.3758/s13421-015-0560-0

Unsworth, N., & Robison, M. K. (2017). Pupillary correlates of covert shifts of attention during working memory maintenance. *Attention, Perception and Psychophysics, 79,* 782–795. https://doi.org/10.3758/s13414-016-1272-7

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

van der Lely, H. K. J., & Stollwerck, L. (1997). Binding theory and grammatical specific language impairment in children. *Cognition, 62,* 245–290. https://doi.org/10.1016/S0010-0277(96)00783-4

Vander Weele, T. (2015). *Explanation in causal inference: Methods for mediation and interaction*. New York, NY: Oxford University Press.

Van Dyke, J. A., Johns, C. L., & Kukona, A. (2014). Low working memory capacity is only spuriously related to poor reading comprehension. *Cognition, 131,* 373–403. https://doi.org/10.1016/j.cognition.2014.01.007

Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language, 55,* 157–166. https://doi.org/10.1016/j.jml.2006.03.007

Van Dyke, J. A., & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language, 65,* 247–263. https://doi.org/10.1016/j.jml.2011.05.002

Wallace, G., & Hammill, D. (1994). *Comprehensive Receptive and Expressive Vocabulary Test–Second Edition*. Austin, TX: Pro-Ed.

Weighall, A., & Altmann, G. T. (2011). The role of working memory and contextual constraints in children's processing of relative clauses. *Journal of Child Language, 38,* 579–605. https://doi.org/10.1017/S0305000910000267

Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in Psychology, 4,* 433. https://doi.org/10.3389/fpsyg.2013.00433

Yuan, S., & Fisher, C. (2009). "Really? She blicked the baby?": Two-year-olds learn combinatorial facts about verbs by listening. *Psychological Science, 20,* 619–626. https://doi.org/10.1111/j.1467-9280.2009.02341.x

Yuan, S., Fisher, C., & Snedeker, J. (2012). Counting the nouns: Simple structural cues to verb meaning. *Child Development, 83,* 1382–1399. https://doi.org/10.1111/j.1467-8624.2012.01783.x