ASHA American Speech-Language-Hearing Association

JSLHR

Research Article

Identifying Linguistic Markers of French-Speaking Teenagers With Developmental Language Disorder: Which Tasks Matter?

Émilie Courteau,[a,b] Guillaume Loignon,[c] Karsten Steinhauer,[b,d] and Phaedra Royle[a,b]

[a] School of Speech-Language Pathology and Audiology, Faculty of Medicine, University of Montréal, Québec, Canada [b] Centre for Research on Brain, Language and Music, Montréal, Québec, Canada [c] Deparment of Education and Pedagogy, Faculty of Educational Sciences, University of Québec at Montréal, Canada [d] School of Communication Sciences and Disorders, Faculty of Medicine, McGill University, Montréal, Québec, Canada

ARTICLE INFO

Article History:
Received October 8, 2021
Revision received April 14, 2022
Accepted September 27, 2022

Editor-in-Chief: Stephen M. Camarata
Editor: Christos Salis

https://doi.org/10.1044/2022_JSLHR-21-00541

ABSTRACT

Purpose: This research aimed to identify reliable tasks discriminating French-speaking adolescents with developmental language disorder (DLD) from their peers with typical language (TL) and to assess which linguistic domains represent areas of particular weakness in DLD. Unlike English, morphosyntax has not been identified as a special area of weakness when compared with lexicosemantics in French preschoolers with DLD. Since there is evidence that subject–verb number agreement is consolidated in later childhood, one might expect morphosyntax to be a particular weakness and marker of French DLD only in (pre)adolescence.
Method: We administered 20 subtasks that assessed linguistic and phonological working memory skills of two groups: 17 adolescents clinically identified as having DLD (M = 14.1 years) and 20 (pre)teens with TL (M = 12.2 years). Using robust statistics that are less affected by outliers, we selected the most discriminating subtasks between our groups, calculated their optimal cutoff score, and derived diagnostic accuracy statistics. We combined these subtasks in a multivariable model to identify which subtasks contributed the most to the identification of DLD.
Results: Seven subtasks were selected as discriminating between our groups, and three showed outstanding diagnostic accuracy: Recalling Sentences, a multi-word task assessing lexicosemantic skills, and a subject–verb number agreement production task. When combined, we found that the latter contributed the most to our multivariable model.
Conclusion: This study provides evidence that the most relevant markers to identify DLD in French teenagers are tasks assessing lexicosemantics and morphosyntactic domains, and that morphosyntax should be considered an important area of weakness in French-speaking teenagers with DLD.
Supplemental Material: https://doi.org/10.23641/asha.21753932

The diagnostic criteria for children with language disorder in the *Diagnostic and Statistical Manual of Mental Disorders–Fifth Edition* (*DSM-5*; American Psychiatric Association, 2013) include early onset of symptoms and persistent difficulties in the acquisition of language caused by comprehension or production deficits. These are characterized by a reduced vocabulary, limited sentence structures, and discourse impairments. Those language deficits are not the result of sensory, or motor impairments or global delays and will result in functional limitations in many areas, including social participation and academic achievement. The new developmental language disorder (DLD) label suggested by Bishop et al. (2017) aligns with the *DSM-5* definition and provides additional guidelines for clinicians (e.g., the first step toward a DLD diagnostic should be to establish functional impairments).

Correspondence to Émilie Courteau: emilie.courteau@umontreal.ca.
Disclosure: *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.
Downloaded from: https://pubs.asha.org Université de Montreal on 01/07/2023, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Both sources converge in saying that language disorders diagnosed at the age of 4 or 5 years usually persist into adulthood. The *DSM-5* specifies that although the language deficit will persist, the specific profile of language strengths and weaknesses is likely to change over the child's development. Therefore, it is essential to investigate which reliable tasks can be used to assess DLD in adolescents. Conti-Ramsden et al. (2001) found that verb production, nonword repetition, and sentence repetition tasks were reliable to identify DLD in 11-year-old preteenagers. However, this still remains to be explored in languages other than English. For Québec French, only a few standardized tests are available to assess DLD, and some of those tests' norms have been shown not to meet psychometric criteria (Bouchard et al., 2009). There is definitely a need for studies that evaluate and establish the best markers for the identification of French teenagers with DLD.

Although both the *DSM-5* (American Psychiatric Association, 2013) and the DLD definition (Bishop et al., 2017) suggest broad criteria for language deficits in DLD, cross-linguistic reviews have revealed weaknesses in grammatical comprehension and production in multiple languages including English, French, Italian, and German (Balilah et al., 2019; Leonard, 2014). Within the domain of grammatical deficits, much attention has been paid to subject–verb agreement as a hallmark of DLD,[1] as evidenced by numerous theories of DLD focused on this feature. Thus, based on French-speaking children, Paradis and Crago (2001) proposed the "extended optional default" account to explain the overuse of finite verb stems (singular present tense forms) in children with DLD, similar to children with typical language (TL) but for an extended period.

In addition to deficits in subject–verb agreement, morphosyntax is thought to be a particular weakness compared with lexicosemantics, which is likely to be a relative strength in DLD. In Ullman and Pierpont's procedural deficit hypothesis (Ullman et al., 2020; Ullman & Pierpont, 2005), (morpho)syntax is believed to be impaired in DLD due to a procedural memory deficit. Language areas subserved by procedural memory are thus expected to be impaired. At the word-form level, the procedural memory supports rule-governed regular morphology processing such as English third-person verb number agreement (e.g., *she sing-s*). At the sentence-level, it underlies agreement—such as number agreement—between sentence constituents (Steinhauer & Ullman, 2002). However, some aspects of lexicosemantics (e.g., word forms and their meaning) are expected to be preserved as the result of a typically developing declarative memory system. This relative strength of lexicosemantics over morphosyntax had also been suggested in studies of English-speaking preschoolers with DLD. For example, Leonard et al. (1999) compared the use of verb morphology of children with and without DLD aged 3;7–5;9 (years;months) with the number of different verbs they used. They found that the use of verb morphology in children with DLD was below expectations for the number of different verbs in their vocabulary, compared with their peers with TL, and suggested that grammatical morphology should be used as a clinical marker for DLD. Interestingly, this has not been robustly observed in French. Elin Thordardottir and Namazi (2007) found that morphosyntactic errors are not a salient feature of spontaneous speech in preschoolers with DLD when compared with lexicosemantic ones. These authors argued that this diverges from the pattern found in English for the same age group. Furthermore, while French-speaking preschoolers with DLD make more errors than their peers with TL on tasks assessing comprehension and production of subject–verb agreement, they were not useful in discriminating preschoolers with and without DLD (Elin Thordardottir et al., 2011). Rather, the combination of a receptive vocabulary task and a nonword repetition task was more sensitive to the disorder in young children (see details below). Since French children with TL are likely to master subject–verb number agreement during school-age years, one could expect it to be a special area of weakness and marker of French DLD only in (pre)adolescence, but this has yet to be studied. Indeed, on the basis of an elicited production task of the irregular verb *faire*, "to do," there is evidence that subject–verb agreement is not automatized until age 8;5 for children with TL (Franck et al., 2004).

This study first aims to describe the linguistic impairments of French teenagers with DLD from Québec, Canada. Through the assessment of participants with 20 tasks often used in clinical and research settings, we intend to evaluate which ones best discriminate between preteens and teens with and without DLD. Our second objective is to identify if morphosyntax is a particular area of weakness when compared with lexicosemantics in French adolescents with DLD.

Linguistic Impairments in Adolescents With DLD, With a Focus on French

Language research on DLD and assessment tasks used by speech-language pathologists (SLPs) usually assume that language has multiple domains, that is, it is multidimensional (Lonigan & Milburn, 2017). For instance, Tomblin and Zhang (2006) give a classic example of how language is assessed through commercially available test batteries: Domains such as grammar and vocabulary will

[1]Note that when we review studies that use the previous common label specific language impairment (SLI), we will use DLD for the sake of clarity, even though we are aware that the diagnostic criteria for these labels are not interchangeable.

Downloaded from: https://pubs.asha.org Université de Montreal on 01/07/2023, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

be evaluated by different tasks in the receptive and expressive modalities, assuming that linguistic domains can be impaired or preserved in an individual (e.g., preserved receptive syntax vs. impaired expressive vocabulary). However, this multidimensionality has been challenged by researchers. Tomblin and Zhang (2006) found that it is valid to consider vocabulary, assessed with word-level tasks, and grammar, reflected by sentence-level tasks, as two dimensions only from the second grade onward (i.e., for students aged 7 or 8 years). While Lonigan and Milburn (2017) found that vocabulary and syntax are two distinct dimensions from preschool onward, their findings suggest that there is still a lot of variance shared by these two linguistic domains. Interestingly, both these studies failed to find evidence supporting the idea that comprehension and production language skills were two different dimensions. Overall, these studies agree on a multidimensional perspective of language at least in older school-age children with and without DLD, offering support for the multidimensionality of adolescents' language abilities. Considering this, we will now review what linguistic domains are likely to be impaired in teenagers with DLD. When studies on French-speaking adolescents with DLD were not available, we first reviewed studies with English-speaking teenagers and then French studies with (pre)adolescents or children.

Morphosyntactic impairments in DLD have been found in many languages and have been accounted for in many frameworks (for reviews, see Balilah et al., 2019; Leonard, 2014), but research has generally focused on preschool or young school-age children. Looking at older children allows us to see which morphosyntactic structures have been acquired and what is still impaired despite many years of practice. Only a few studies have detailed the morphosyntactic skills of older French-speaking children with DLD. Prigent et al. (2015) examined the use of complex morphosyntactic structures in spontaneous speech by a group of French-speaking DLD (pre)teens aged 11;6, on average. They showed that participants with DLD produced fewer complex morphosyntactic structures when compared with a control group aged 7;8, on average, and matched on morphosyntactic comprehension skills with the DLD group, suggesting that avoidance of complex morphosyntactic structures is characteristic of French-speaking (pre)teens with DLD. The less frequent use of complex structures in spontaneous speech was also observed for syntax. Tuller et al. (2012) showed that French teenagers with DLD produced significantly more ungrammatical utterances compared with three groups of younger children with TL aged 6, 8, and 11 years: 15.5% in the DLD group compared with less than 5% in all TL groups. Crucially, however, the DLD group also avoided producing complex sentences. Turning to elicited production tasks, Rose and Royle (1999) used a sentence completion task to elicit the production of 12 French verbs in the present or past tense by 20 participants from families with DLD aged 9–46 years. They found deficits in tense production in the DLD group when compared with eight controls matched on age and educational background. Using a completion task for subject–verb number agreement, Franck et al. (2004) elicited the production of the irregular verb *faire*, "to do" (*il fait* [fɛ]–*ils font* [fɔ̃], "he does–they do"). Based on 60 children with TL aged 5–8;5 and eight children with DLD aged 5–9;4, they found that children with TL aged 7 years were still producing 25% subject–verb agreement errors, whereas it dropped to 5.4% by age 8;5. Participants with DLD (aged 8;8 on average, no individual data are presented), still produced errors 20% of the time. Since these two studies were based on participants with DLD within a wide age range and targeted only a small number of irregular verbs, mastery of subject–verb agreement by older French children and teenagers with and without DLD remains to be described.

Lexicosemantic skills (i.e., vocabulary) has not usually been thought to be a marker for DLD in school-age children, at least in English. For instance, although Conti-Ramsden et al. (2001) administered vocabulary tasks to their English-speaking participants, these were not considered potential positive psycholinguistic markers of DLD and, thus, were not included in their analyses of diagnostic accuracy. However, McGregor et al. (2013) assessed 177 DLD and 325 TL English-speaking children and teenagers in Grades 2, 4, 8, and 10 on their vocabulary's breadth, through the number of words defined correctly, and vocabulary's depth, measured as the quantity of correct information in each definition. Participants with DLD showed deficits on both measures throughout all age groups. Impairments were also found on receptive vocabulary. Using the Peabody Picture Vocabulary Test (PPVT), Rice and Hoffman (2015) tested DLD and TL children, teens, and young adults in a longitudinal study from ages 2;6 to 21 years and found lower performance for participants with DLD across the duration of the study. Targeting French-speaking children, Elin Thordardottir et al. (2011) found that the *Échelle de vocabulaire en images Peabody* (EVIP; the French version of the PPVT; Dunn et al., 1993) had a relatively high level of sensitivity and specificity in identifying 5-year-old children with DLD.

We have reviewed aspects of morphosyntax and lexicosemantics that appear to be impaired in older children and teenagers when assessed with production or comprehension tasks. One way of assessing these two linguistic domains simultaneously is with sentence repetition tasks. Based on French-speaking 10-year-old children with and without DLD, Leclercq et al.'s (2014) study suggests that the ability to repeat sentences accurately is subserved by two factors: a morphosyntactic factor and a lexical one.

Downloaded from: https://pubs.asha.org Université de Montreal on 01/07/2023, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Using a principal component analysis, their results showed that both factors contributed almost equally to a score on the sentence repetition task: 52.56% of the variance was explained by the morphosyntactic subscore, and 43.92% was explained by the lexical subscore. This kind of task has shown excellent discriminating accuracy for identifying DLD in many languages, including English (e.g., Conti-Ramsden et al., 2001) and French (e.g., Leclercq et al., 2014; Tuller et al., 2018). Tuller et al. (2018) used a sentence repetition task specifically designed to address the complex syntax deficits in DLD to discriminate between French-speaking children with and without DLD between the ages of 5 and 8 years. They argue that their task was efficient because it targets complex French syntax structures, which have been shown to be impaired in DLD (e.g., Delage & Frauenfelder, 2020).

Adolescents' morphosyntax and lexicosemantic ability has also been assessed using grammaticality judgment tasks, a method used in DLD research but rarely in clinical settings. Miller et al. (2008) showed that English-speaking 16-year-olds with DLD were significantly less successful at identifying sentences containing subject–verb agreement omission and commission errors, compared with their peers with TL matched on age. Similar findings were reported by Noonan et al. (2014) for 8-year-olds with DLD and by Haebig et al. (2017) in 15-year-olds with DLD. Furthermore, the latter authors found that the DLD group performed more poorly on a judgment task containing lexical–semantic errors on verbs when compared with the TL group. Lower performance on judgment tasks was also found in French-speaking children with DLD. In a case study of a French-speaking child with DLD aged 8 years, Poulin et al. (2015) found that, in an oral task with visual support, his ability to identify gender agreement errors on adjectives, but not semantic errors, was impaired in comparison to age-matched children with TL, but not to younger ones (6 years of age). Maillart and Schelstraete (2005) observed reduced ability to detect sentences containing agreement or tense marking errors in a group of 9-year-old French children with DLD compared with a TL group matched on receptive grammatical skills, aged 5;4, while they performed similarly on syntactic word order errors. Rose and Royle (1999) found that 20 (pre)teens and adults (aged 9–46 years, from French-speaking families with DLD) performed worse than the control TL group on identification of determiners, prepositions, verb tense, and number agreement or argument–structure errors (e.g., missing complements) in sentences. Using a lexical decision task in which French participants had to identify if a heard word was a pseudoword or not, Quémart and Maillart (2016) found lower performance for the 10-year-old DLD group when compared with children with TL matched on aged or receptive vocabulary.

Another area that has been shown to be impaired in DLD is phonological working memory[2] (Leonard, 2014). Using forward and backward digit span tasks, Arslan et al. (2020) found impaired phonological working memory skills in two groups of French-speaking DLD children aged 7–11 years and teenagers aged 12–18 years when compared with age-matched control groups. Interestingly, there was no difference between the teen groups on visuospatial working memory skills assessed through the forward and backward Corsi (1972) blocks test, but the younger DLD group showed significantly lower performance than their age-matched peers with TL on the backward Corsi block span, suggesting that these skills can normalize with age. Overall, these studies reveal that teenagers with DLD are likely to experience linguistic deficits expressed by impairments in morphosyntax, lexicosemantics, and phonological working memory.

Linguistic Impairments: From Research to Diagnosis

The vast majority of the findings we reviewed in this section were interpreted as signaling linguistic impairments in DLD based on group performance. In research settings, a task will be administered to a group of individuals known to have DLD and a control group with TL, and if the mean difference between groups is significantly different, researchers will conclude that there is a deficit in the DLD group on the construct(s) or linguistic domains evaluated by the task. However, in clinical settings, practitioners need to be able to determine if a single individual has a deficit based on their performance on a given screener or evaluation tool. To do so, usually the practitioner will use a task where a cutoff is provided (i.e., a threshold score; Hajian-Tilaki, 2013). If the individual's task score is below the threshold, is it likely that they present with a DLD. As we know, no task is perfect, and the diagnostic accuracy of a task can be described using multiple measures. First, task sensitivity reflects the true positive rate, which represents the proportion of participants with a documented DLD who are identified as such by a given test. Second, specificity reflects the true negative rate, the proportion of participants with TL development who are identified as such. For these two measures, a proportion above 90% is considered valid and good, a proportion between 80% and 89% is considered fair, and a

[2]For the purposes of this article, we will use the term *working memory* in the sense of "a limited capacity system allowing the temporary storage and manipulation of information" as defined by Baddeley (2000, p. 418). This system includes a phonological and a visuospatial component, which we will refer to as the phonological and visuospatial working memory. Phonological working memory is also referred to as the phonological short-term memory, as by Leonard (2014), for example, when he lists weaknesses related to DLD across languages.

Downloaded from: https://pubs.asha.org Université de Montreal on 01/07/2023, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

proportion below 80% is considered unacceptably low (Plante & Vance, 1994). These two measures can be combined in an index: the likelihood ratio. A positive likelihood is the ratio of true positives to false positives; higher values indicate more informative tests. A ratio of 10 is considered strong (Jaeschke et al., 1994) and indicates that the likelihood of having a DLD would be 10 times higher if the participant's score is below the threshold than if it was above. Inversely, a negative likelihood is the ratio of false negatives to true negatives; values close to 0 indicate more informative tests: A ratio of .1 is considered strong. Other qualitative terms used to describe positive/negative likelihood ratios are the following: 5–10/.1–.2, moderate; 2–5/.2–.5, small and sometimes important; 1–2/.5–1, small but rarely important.

Lastly, the quality of a task can also be described with a receiver operating characteristic (ROC) curve. This curve plays a major role in judging the diagnostic accuracy of tasks to differentiate the true state of participants (DLD or TL) and to find the optimal cutoff values (Hajian-Tilaki, 2013). The ROC curve is a plot of the true positive rate (sensitivity) to false positive rate (1 − specificity) for all possible cutoff scores. Using this curve will let us choose the optimal cutoff score or threshold for a task. To do so, the derived area under the curve (AUC) measure is used, which can be interpreted as such: .5 is no better than chance, .5–.7 equals poor discrimination, .7–.8 is acceptable, .8–.9 is excellent, and over .9 is considered outstanding classification. The assessment of linguistic deficits is a challenge for clinicians in Québec given the sometimes-low diagnostic accuracy of adapted tests and limited number of standardized tools available in French for children and adolescents, a situation we describe in the next section.

Language Assessment of Québec French–Speaking Adolescents With DLD

When assessing the language of younger children who are suspected of having language impairment, one of the key outcomes is to establish a diagnosis of DLD based on functional and language impairments, as proposed by Bishop et al. (2017). Since adolescents with DLD already have been given a diagnosis, language evaluation will have other main goals. In clinical practice, the results of these evaluations are often used to determine which linguistic domains or structures SLP therapy should focus on. Another objective of assessing language in teenagers with DLD, at least in Québec, is to establish language disorder severity. Even if it is not recommended by Québec and international standards (Bishop et al., 2017; Tessier et al., 2017), low scores on standardized tests to confirm a disorder's severity is still a widespread criterion used in Québec to access services (Breault et al., 2019). A challenge is that only few standardized tests are available for the Québec French teenager population: in their review of oral language tests. Monetta et al. (2016) listed five that could be used with an adolescent population, and only one was normed and validated for the French Québec's population, that is, the *Évaluation clinique des notions langagières fondamentales–version pour francophones du Canada* (CELF CDN-F; Secord et al., 2009), standardized up to 16 years of age. As a result, SLPs from Québec have the option to use adapted (or even less ideal, translated) standardized English tests or French tests standardized in France. These tests are rarely based on appropriate cultural and linguistic norms for the Québec population (Bouchard et al., 2009). In addition to clear lexical differences, grammatical ones might also emerge. For instance, Courteau et al. (2019) showed that adult speakers of Québec French did not systematically process incorrect omission of subject–verb plural liaison, while Bourget (1987) suggests that this kind of liaison is rarely produced orally in informal Québec French. Speakers would very rarely use the plural feminine third-person pronoun *elles* [ɛlz], "they.fem." Instead, it is replaced by *ils* [ɪlz], "they.MASC," which undergoes /l/ deletion (i.e., *ils* pronounced [i], [iz], or [j], but rarely [ɪl/z], the standard forms for plural). Tasks based on different language varieties might thus underevaluate linguistic abilities in Québec French speakers.

Furthermore, many tests are not based on appropriate norms that meet psychometric criteria for the target population (Bouchard et al., 2009). Indeed, Elin Thordardottir et al. (2011) showed that among 78 monolingual speakers of Québec French, mean group scores were 1 *SD* higher than the published norms of the French version of the PPVT (EVIP; Dunn et al., 1993). This could be attributed to the fact that the published norms were based on pan-Canadian francophones that included monolingual, but also bilingual, French speakers. This leads to the underestimation of language difficulties of monolingual Québec French–speaking children (Godard & Labelle, 1995, cited in Elin Thordardottir et al., 2011). Considering that 82% of the 8 million inhabitants in Québec have French as their first spoken language (Statistics Canada, 2016), there is an urgent need for research on linguistic markers of Québec and Canadian French-speaking teenagers with DLD.

This Study

We demonstrated in the previous sections that there is an urgent need for studies that identify reliable tasks that discriminate French-speaking adolescents with and without DLD—and even more so for the Québec population. In order to obtain a more comprehensive portrait of this population, research should combine typical tasks used in the two contexts where we find language assessments of

Downloaded from: https://pubs.asha.org Université de Montreal on 01/07/2023, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

teenagers with DLD, that is, clinical and research settings. For instance, it has been shown, mostly in English, that teenagers with DLD are impaired in their grammaticality judgment abilities, but to our knowledge, the diagnostic accuracy of this type of task has never been compared with tasks used in clinical settings, such as sentence repetition. The first objective of this study is to examine the diagnostic accuracy of tasks used in research and clinical setting to discriminate between a group of French-speaking teenagers known to have DLD since childhood and a group of preteens and teens with TL. The tasks selected for this study were taken from published and experimental materials with the goal to cover language subdomains that have been identified as weaknesses in older children and adolescents with DLD. We examined which of these tasks provided the highest degree of accuracy in identifying adolescents with DLD. Studies on older children and adolescents with DLD tend to show heterogeneous patterns of severity as a population and within individuals (Conti-Ramsden, 2008), with severity differences depending on the linguistic domains or selectively impaired domains (Friedmann & Novogrodsky, 2009). Considering this and the lack of studies on French adolescents with DLD, we do not have clear expectations as to which tasks are expected to be the most discriminating or difficult for our population. Elin Thordardottir et al. (2011) showed that, among 10 language tasks, five accurately discriminated between French-speaking 5-year-olds with and without DLD: receptive vocabulary, receptive morphosyntax skills, nonword and sentence repetition, and a following-directions task. Closer to the age of our group, the sentence repetition task has already been demonstrated to be discriminating in older French-speaking children with DLD (Leclercq et al., 2014), and significant differences between groups of older children with and without DLD were found in a task eliciting the production of subject–verb number agreement for the irregular verb *faire*, "to do" (Franck et al., 2004). A second objective of this study was to assess linguistic domains that especially represent an area of weakness in French adolescents with DLD. To do so, we compared the tasks used in this study that yielded the best discriminating accuracy for teens with DLD, and investigated which one contributed most to the identification of DLD in a multivariable model. In line with research on DLD in many languages, deficits in morphosyntax have been found in French DLD (e.g., Rose & Royle, 1999; Stanford et al., 2019). In contrast to English (e.g., Leonard et al., 1999), morphosyntactic difficulties have not been identified as a particular weakness in French-speaking preschool children with DLD when compared with lexicosemantic skills (Elin Thordardottir et al., 2011; Elin Thordardottir & Namazi, 2007). However, it may be that the deficits in morphosyntax are characteristic of French DLD only later, at adolescence, since children with TL are likely to have mastered subject–verb number agreement by the age of 8.5 years (Franck et al., 2004).

Method

Participants

Thirty-seven French-speaking children and teenagers participated in this study, which is part of a larger research project on neurocognitive processing in DLD[3] (see Courteau et al., 2019). The protocol was approved by the University of Montréal Research Ethics Board for educational and psychology research (CERES-15-070-D). In accordance with the Declaration of Helsinki, all participants' parents gave written consent for their child's participation, and the participants themselves gave oral consent prior to the first experimental session. All had a hearing screening on the first day of assessment (500–8000 Hz at 25 dB in at least one ear). Their mother tongue was French and was their language of instruction and daily use.

A group of 17 teenagers with DLD (DLD group), including 10 girls, aged between 12 and 15 years (M = 14.1, SD = 0.72), served as the clinical group, with which the language measures' diagnostic accuracy was determined. The majority (n = 14) were recruited from a specialized private school for children and adolescents with learning disabilities in Montréal (Québec, Canada) through a letter of invitation sent by the school SLP to the parents of students meeting the selection criteria. It should be noted that this school excludes children with disruptive behavior, possibly explaining why our group of participants with DLD includes more girls than boys. The other participants were recruited from a parent's association for children with DLD.

The inclusion criteria of DLD were a clinical diagnosis of DLD since childhood, functional impairments that meet the DLD definition as detailed by Bishop et al. (2017), and persistent language impairments. All participants had a documented history of DLD and a complete SLP language evaluation (including discourse and pragmatic domains) resulting in a diagnosis.[4] All teenagers of the DLD group had been diagnosed before kindergarten or during the first year of primary school and maintained significant functional impairments needing adaptations to succeed in school. These were, for the most part,

[3]The participants are the same as in the studies of Courteau et al. (under revision) and Pourquié et al. (under revision); thus, their description and characteristics are similar in content and wording across these articles.

[4]Following their diagnostic, all participants with DLD received individual speech-language therapy, the duration, intensity, and therapeutic goals of which varied. Sixteen of the 17 participants with DLD took or take part in speech and language group therapy.

Downloaded from: https://pubs.asha.org Université de Montreal on 01/07/2023, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

accommodations in regular classes or enrolment in special ones. Note that many participants had comorbid disorders, such as attention-deficit/hyperactivity disorder (ADHD) and dyspraxia. These disorders do not preclude a DLD diagnosis (see Statement 9; Bishop et al., 2017). A study by Redmond et al. (2015) showed that ADHD comorbidity with DLD—and TL development—does not increase children's errors on language assessment tasks such as sentence recall. However, the dominant clinical profile of these participants was the presence of persistent language difficulties. Exclusionary criteria were the presence of associated biomedical conditions, such as intellectual disabilities and autism spectrum disorder, and bilingual language acquisition (i.e., being exposed from birth or before 5 years of age to French and another language simultaneously).

Twenty participants with no history of language impairment (seven girls), aged between 7 and 14 years (M = 12.2, SD = 2.25) formed the TL group. Their typical developmental status was established via a questionnaire filled out during an interview with their parents. None had any significant prenatal or perinatal complications, extended hospitalization, or serious illness.

Both groups were matched on nonverbal abilities, that is, visuoattentional working memory, using four tasks within the Cognitive Experiments IV v2 pack of the Presentation software (Version 18.0, Neurobehavioral Systems, Inc., http://www.neurobs.com). Visuospatial working memory was assessed with the forward and backward Corsi blocks tasks (Corsi, 1972) and by a delayed match-to-sample (DMTS) task of nonverbal stimuli (Daniel et al., 2016) with delays of 1 or 5 s. Participant characteristics for both groups are presented in Table 1. To compare groups statistically, we used Brunner–Munzel tests (Brunner & Munzel, 2000), as recommended by Rietveld and van Hout (2015) for skewed data with small sample sizes. The Brunner–Munzel is a robust nonparametric test that checks the stochastic superiority of a group, expressed by the Brunner–Munzel statistic (t_{bm}), a p value, and a common-language effect size (CLES), indicating in our case the probability of a random observation from the TL group being larger than a random observation from the DLD group, with .5 being at chance. Differences between groups were found in age (DLD > TL) and schooling (DLD > TL). Note that participants were meant to be matched on the age variable, but that recruitment was halted due to the COVID pandemic. Since the DLD group is significantly older than the TL group, higher TL group scores cannot be attributed to age. See Figure 1 for a display of the age variable distribution.

Procedure

Experimental sessions took place in a quiet room either at the participants' high school or at the University of Montréal in the fourth author's lab. Participants were individually tested for two 2- to 2.5-hr sessions where, in the first hour, they participated in an event-related potential (ERP) experiment (Courteau et al., under revision). Testing was conducted by a Québec-accredited SLP (i.e., the first author) or trained research assistants. All experimenters had French as their native language. The tasks used in this study can be classified into two categories and will be briefly described in the following order: (a) those commonly used by SLPs in clinical settings to assess the language skills of Québec French adolescents and (b) those used in DLD research. See Supplemental Material S1 for a detailed presentation of tasks.

Three tasks were selected from the CELF CDN-F French version standardized for Canadian French speakers aged 4–16 years old (Secord et al., 2009) and were administered as recommended by the manual. We used the Recalling Sentences task, where participants repeated orally presented sentences, and the Word Classes task, which assessed the ability to understand lexicosemantic relationships between orally presented words by choosing the two words that go together (receptive subtask) and explaining this relationship (expressive subtask). Third, the number repetition tasks consisting of a forward and

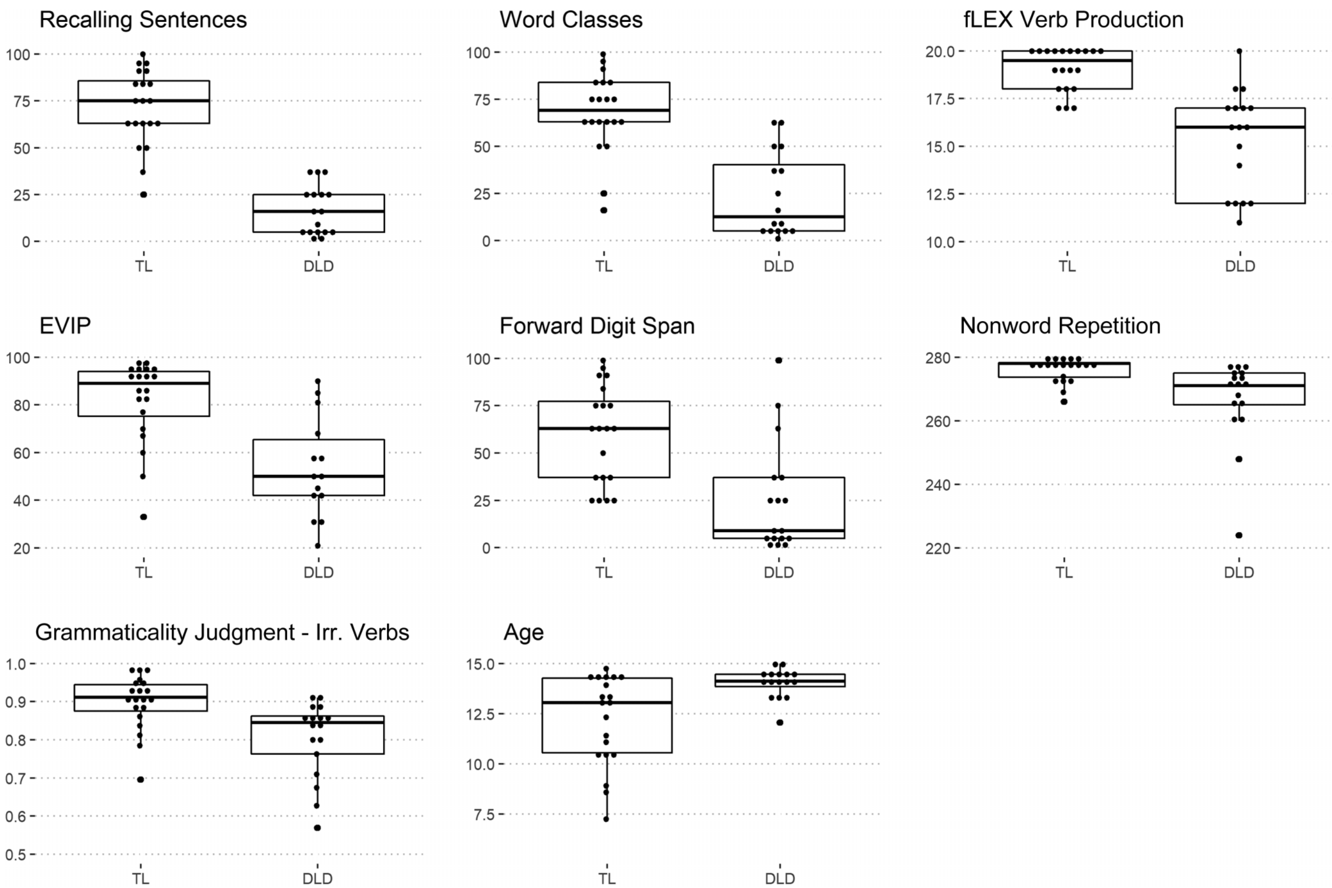
Table 1. Participant characteristics with comparisons between groups made by the Brunner–Munzel test.

	TL group (n = 19)		DLD group (n = 17)		Brunner–Munzel tests		
Variable	***M***	***SD***	***M***	***SD***	t_{bm}	***p***	**CLES**
Age	12.21	2.25	14.09	0.77	3.39	.002	0.24
School	5.9	2.20	7.53	0.51	2.30	.03	0.30
Corsi–F	5.55	1.76	5.56	1.55	0.12	.91	0.51
Corsi–B	5.60	1.76	4.94	1.06	1.18	.25	0.61
DMTS–1 s	0.89	0.11	0.88	0.10	0.48	.63	0.55
DMTS–5 s	0.82	0.15	0.84	0.13	0.24	.81	0.47

Note. Chronological age (Age) and schooling (School) are expressed in years. Corsi block scores reflect forward (Corsi–F) and backward (Corsi–B) untransformed spans, and delayed match-to-samples represent percentage accuracy for 1-s (DMTS–1 s) and 5-s (DMTS–5 s) delays. TL = typical language; DLD = developmental language disorder; CLES = common-language effect size.

Downloaded from: https://pubs.asha.org Université de Montreal on 01/07/2023, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Figure 1. Results for both DLD and TL groups on seven discriminating subtasks and age. The *y*-axis indicates the score. Scores are presented as percentile scores except for fLEX subject–verb agreement production (max: 20 target verbs), nonword repetition (max: 280 target phonemes), grammaticality judgment of subregular and irregular verbs (*A*-score: 1 being perfect, .5 indicating chance). The age scale is expressed in years. DLD = developmental language disorder; TL = typical language; EVIP = *Échelle de vocabulaire en images Peabody*; Irr. = irregular.



backward digit span were used to evaluate phonological working memory. We chose the EVIP (a standardized Canadian French adaptation of the PPVT; Dunn et al., 1993) to evaluate receptive vocabulary.

Expressive vocabulary was assessed with an action (verb) naming task from the French version of the experimental fLEX test (see Pourquié, 2017, for details). We evaluated subject–verb number agreement skills with two tasks from fLEX, which included 20 verbs[5] that had an audible word final agreement number cue and occasionally a vowel change (e.g., *il rugit* [ilʁyʒi], "he roars," vs. *ils rugissent* [ilʁyʒɪs], "they roar"). The expressive task assessed subject–verb number agreement on verbs via elicited sentence production, and the receptive task assessed their understanding. An interesting feature of the fLEX test is that it allows one to test both lexical and morphosyntactic aspects of language associated with verb production and comprehension, as the same verbs are used in the three tasks (expressive vocabulary, production, and comprehension of subject–verb number agreement). The fLEX test is a research tool and does not provide norms, but we selected it for its thorough control of linguistic cues. We used the nonstandardized French Québec nonword repetition Courcy task (Elin Thordardottir et al., 2011, Appendix) to evaluate phonological working memory. This task consists of 40 items; the total of correctly repeated phonemes, with a maximum of 180, was scored.

As tasks commonly used in research on language acquisition, we first used two grammaticality judgment tasks based on an alien-learning paradigm (Courteau et al., 2013) where an alien practices French and sometimes makes mistakes. Participants listened to prerecorded sentences while looking at pictures and judged if sentences were correct or not. Participants' grammaticality judgments were quantified with *A*-scores, a bias-adjusted measure of sensitivity that includes the participant's ability to correctly classify presented sentences as containing an

[5]Verbs were either subregular or irregular verbs from the second and third conjugation groups.

Downloaded from: https://pubs.asha.org Université de Montreal on 01/07/2023, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Table 2. Linguistic and working memory subdomains assessed in this study with their associated subtasks.

Subdomains	Subtasks
Lexicosemantics	Word Classes, EVIP, fLEX action (verb) naming, lexical–semantic judgments for nouns and verbs
Morphosyntax	fLEX subject–verb agreement production and comprehension, grammaticality judgments for gender agreement in noun phrases, and subject–verb number agreement in sentences containing subregular or irregular verbs.
Phonological working memory	Forward and backward digit span, nonword repetition
Visuospatial working memory	Forward and backward Corsi blocks, DMTS–1 s and DMTS–5 s
Lexicosemantics, morphosyntax, and phonological working memory	Recalling Sentences

Note. EVIP = *Échelle de vocabulaire en images Peabody;* DMTS–1 s = delayed match-to-sample, percentage accuracy for 1-s delay; DMTS–5 s = delayed match-to-sample, percentage accuracy for 5-s delay.

error or not (A'-score, corrected version; Zhang & Mueller, 2005). A-scores of 1 reflect perfect discrimination and .5 chance levels. In the first task, adapted from Poulin et al. (2015), participants looked at pictures while listening to 16 sentences that were either correct (n = 4) or contained errors targeting the noun phrase (n = 12). Errors included auditory–visual lexicosemantic mismatches on nouns (n = 4) and morphosyntactic gender agreement errors on determiners or adjectives (four each). The second task was run during an ERP experiment (Courteau et al., under revision). Each participant heard 300 spoken grammatical sentences paired with a picture that matched or mismatched its morphosyntactic (n = 240) or semantic (n = 60) features: In each condition, half of the visuoauditory pairs were mismatches. They judged if the pairs were a match or not by pressing a key. Lexicosemantic errors were created using a verb that did not match the depicted action (n = 30). Subject–verb number agreement errors were created by varying the number of visually presented agents and morphosyntactic number cues in the auditory stimuli. This was done using either irregular and subregular verbs with an audible ending (n = 60; e.g., visual: [A LION ROARS], auditory: *En soirée, ils *rugissent* [ilʁyʒɪs] *dans la savane*, "In the evening, they *roar in the jungle") or with regular agreement morphophonology, where the plural number cue "s" [z] was realized through liaison between the plural pronoun form and a vowel-initial verb (n = 60). We also assessed visuospatial working memory (see description in the Participants section). Interrater reliability was calculated on all tasks that involved a verbal response for every participant with DLD and four of the participants with TL. Based on Krippendorff (2004), interrater reliability percentage of agreement was at a minimum of 95% and a maximum of 100% across subtasks (Duquette et al., 2020).

In total, we administered 20 subtasks, which generated 26 scores per participant. Three score types were produced depending on the task: 13 raw scores corresponding to the untransformed total number of subtask items successfully completed by the participant, seven percentile rank scores derived from the same-age norm group for standardized subtasks, and six A-scores for the grammaticality judgment subtasks. Note that one participant did not complete the Word Classes and Corsi block tasks (DLD-01, 22/26 scores), and another did not complete the DMTS tasks (TL-06, 24/26 scores). We listed, in Table 2, to which linguistic or working memory subdomains[6] each subtask corresponded according to the tests' manual, when available, or based on the literature as presented in the introduction (for a detailed list of score types and subdomains associated with each subtask, see Supplemental Material S1, Table S1).

Analyses

Selection of Subtasks

First, we applied a variable selection procedure to identify which of the 26 subtask scores had the potential to discriminate between groups and to avoid multicollinearity problems, a common concern with multivariable models. Using RStudio Version 1.4.1103-4 (RStudio Team, 2020), we calculated the information gain (IG; Azhagusundari & Thanamani, 2013) for each variable and rejected those that had a null IG. Next, based on test specifications and IG, we eliminated variables that reflected pairs of scores that originated from the same subtask, such as raw and percentile equivalents of the same measurement, retaining the score with the best IG.

[6]Grammaticality judgments involve lexical–semantic and morphosyntactic, but also metalinguistic, knowledge. Two anonymous reviewers made the point that the contribution of metalinguistics to this type of task is, however, debated. We agree. However, metalinguistic skills are also required in other tasks that we did not associate with this domain on the basis of the test documentation, such as the CELF CDN-F Word Classes production subtask. Since the underlying role of metalinguistics as a cognitive domain in the tasks used in the assessment of DLD is beyond the scope of this article, we abstain from considering metalinguistics as an assessment domain.

Downloaded from: https://pubs.asha.org Université de Montreal on 01/07/2023, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Group Comparisons

Multiple comparisons using Brunner–Munzel tests (Brunner & Munzel, 2000) implemented in the *bunnermunzel* R package (Hui et al., 2020) were applied to assess the difference between TL and DLD groups. In our case, this test estimates the probability that a participant randomly drawn from the TL group will have a higher score than a participant randomly drawn from the DLD group. We applied a Bonferroni–Holm adjustment to resulting *p* values to control the false discovery rates (Abdi, 2010). We report the resulting effect size with 95% confidence intervals (CIs).

Diagnostic Accuracy of Subtasks

The subtasks were considered as tests with threshold scores, that is, cutoff scores, and two possible outcomes: Below the cutoff scores, the participant is assumed to have DLD, and above the cutoff scores, the participant is assumed to have TL development. To analyze the discriminatory ability of the subtask's scores, we identified optimal cutoff scores based on our sample data and calculated measures of diagnostic accuracy. The optimal cutoff scores were estimated by a bootstrap procedure that randomly resampled our groups but with replacements, a thousand times, using the *multi_cutpointr* function of the R package *cutpointr* (Thiele & Hirschfeld, 2020). The selected cutoff point for each variable was the point that maximized the sum of sensitivity and specificity. To mitigate sample bias, the cutoff point was recalculated as the midpoint between the optimal cutoff point and the next lowest score. Ties were resolved by returning the mean of conflicting cutoff points. Note that the recommended cutoff scores for subtasks from standardized tests, generally start at −1 *SD*, or about the 16th percentile (Conti-Ramsden et al., 2001). We report several measures of diagnostic accuracy including sensitivity, specificity, positive and negative likelihood, and AUC. Details on these measures can be found in the Language Assessment of Québec French–Speaking Adolescents With DLD section. For all measures associated with the optimal cutoff scores, we calculated their 95% CI, representing percentiles 2.5 and 97.5, based on the distribution produced by 1,000 bootstrap iterations.

Combination of Subtasks

To identify which subtasks contributed the most to the identification of DLD, we fitted a multivariable logistic regression model to predict the group. The predictor variables were the selection of subtasks previously identified as discriminating between TL and DLD. Collinearity of the remaining six predictor variables was first assessed using the *findLinearCombos* function from R's *caret* package (Kuhn, 2009), which did not reveal multicollinearity problems. The same method applied to centered and normalized scores also did not reveal multicollinearity problems. Since a regular logistic regression procedure resulted in near-perfect separation that prevented us from producing the relevant statistics, we used a regularized logistic regression procedure R's *glmnet* package (Hastie et al., 2016). This method required that all the variables be on the same scale prior to model fitting. All subtasks with nonpercentile scores were centered and scaled, by subtracting the mean and dividing it by the standard deviation. Subtasks with percentile scores were converted to *z* scores using the normal distribution. The lambda-regularization parameter, used to determine how strict the regularization is, was set by a leave-one-out cross-validation procedure. We selected the smallest lambda value that minimized cross-validated classification errors. The relative contribution of the variables to the model was estimated with the permutation method (Altmann et al., 2010) implemented in the *vip* R package (Greenwell & Boehmke, 2020). This permutation method measures the difference to a performance metric when the values of a predictor variable are shuffled, in our case the AUC when a subtask's scores are shuffled across all participants, thus making this variable uninformative. We report the resulting coefficients across 100 repetitions of the permutation procedure to rule out accidental patterns in the shuffled data, along with the relative importance of the variables.

Results

Selection of Subtasks

Out of 26 possible scores, 17 had a null IG (IGs for all subtask scores are presented in Supplemental Material S1, Table S2). Of the remaining nine scores, two pairs of scores stemmed from identical or similar subtasks. Both the Recalling Sentences task's percentile score (IG = 0.52) and raw score (IG = 0.27) had a positive IG, so the score with the lower IG was removed from the selection. The Word Classes production subtask score was removed, given that it is dependent on the Word Classes receptive score, which was kept in the selection because of its higher IG. The Digit Span total score was removed because it is the scaled sum of the forward and backward scores from the same task. The final variable set was composed of four age-based percentile scores (Recalling Sentences, receptive Word Classes,[7] Forward Digit Span, and EVIP), two raw scores (fLEX subject–verb production: maximum 20 verbs; nonword repetition: maximum 280 phonemes), and one *A-score* (grammaticality judgment on

[7]Note that for the sake of simplicity, we use the subtask name Word Classes when referring to the receptive Word Classes subtask in the rest of the article.

Downloaded from: https://pubs.asha.org Université de Montreal on 01/07/2023, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Table 3. Group comparisons on seven discriminating subtasks.

Subtasks	Group mean (SD)		Brunner–Munzel tests		
	TL	DLD	t_{bm}	p	CLES [CI]
Recalling Sentences	71.3 (20.3)	16.5 (13)	–28.88	< .001	0.02 [–0.02, 0.05]
Word Classes	67.8 (21.2)	24 (22.4)	–10.8	< .001	0.07 [–0.01, 0.15]
fLEX verb production	19 (1.1)	15.3 (2.7)	–8.27	< .001	0.09 [–0.01, 0.19]
EVIP	81.8 (17.5)	46.3 (25.1)	–7.27	< .001	0.11 [0.01, 0.22]
Forward Digit Span	59.9 (25.7)	25.2 (29)	–4.34	< .001	0.18 [0.02, 0.33]
Nonword repetition	276.2 (3.8)	266.7 (13.4)	–5.61	< .001	0.15 [0.03, 0.28]
Grammaticality judgment	0.9 (0.1)	0.8 (0.1)	–4.64	< .001	0.18 [0.04, 0.32]

Note. Grammaticality judgment on subject–verb number agreement. Brunner–Munzel *p* values are presented with Bonferroni–Holm adjustments and common-language effect sizes (CLES) with their 95% confidence intervals (CI). TL = typical language; DLD = developmental language disorder; EVIP = *Échelle de vocabulaire en images Peabody*; Irr. = irregular.

subject–verb number agreement). Figure 1 shows the distribution of the final selection of scores.

Group Comparisons

We examined group differences on the seven discriminating tasks as shown in Table 3. The TL group showed significantly better performance in all subtasks as seen by their higher group means and as demonstrated by *p* values below .001. Three subtasks were found to be the most discriminating with a CLES below 0.10: Recalling Sentences, Word Classes, and fLEX subject–verb agreement production. The other tasks had CLES between 0.11 and 0.18.

Diagnostic Accuracy of Subtasks

We identified optimal cutoff scores and related measures of diagnostic accuracy, as shown in Table 4. Recall that four of our seven more discriminating subtask scores are age-based percentile ranks. The recommended cutoff score for standardized tests is typically −1 *SD*, or about the 16th percentile. Results showed that the optimal cutoff score for the Forward Digit Span task was at the 17th percentile, with the value 16 being part of the CI, and thus similar to what is recommended by the test (CELF CDN-F; Secord et al., 2009). We found much higher cutoff scores for Recalling Sentences (43.5th percentile), receptive Word Classes (62.5th percentile), and the EVIP (59th percentile). For these subtasks, the CIs did not include the 16th percentile, indicating that a cutoff score of 16 is unlikely based on our sample.

Related to our optimal cutoff scores, Recalling Sentences showed the highest sensitivity and specificity, above .90, followed by Word Classes with good to fair sensitivity and specificity, respectively .94 and .8, and fLEX with fair levels on both measures, .82 and .85. All other tasks exhibited measures under .80 on either one of these measures, indicating unacceptably low accuracy. Regarding likelihood ratios, only Recalling Sentences, Word Classes, and fLEX subtasks revealed strong, moderate, or near-moderate effects on both positive and negative ratios. It is not surprising that these subtasks also revealed the highest AUCs, of over .9 or .8, when considering their CIs, reaching outstanding or at least excellent classification of

Table 4. Bootstrap estimated optimal cutoff scores and their derived measures of diagnostic accuracy.

Subtask	Optimal cutoff	Sensitivity	Specificity	PosLH	NegLH	AUC
Recall.	43.5 [31, 50]	1 [.92, 1]	.9 [.76, 1]	10 [4.25, ∞]	0[a] [0, 0.08]	.98 [.94, 1]
Word C.	62.5 [20.5, 62.75]	.94 [.68, 1]	.8 [.64, 1]	4.69 [2.57, ∞]	0.08 [0, 0.34]	.93 [.82, .99]
fLEX	17.5 [16.5, 18.5]	.82 [.57, 1]	.85 [.64, 1]	5.49 [2.54, ∞]	0.21 [0, 0.43]	.91 [.8, .99]
EVIP	59 [47.5, 90.5]	.76 [.61, 1]	.9 [.55, 1]	7.65 [2.14, ∞]	0.26 [0, 0.42]	.89 [.77, .97]
F. Digit	17 [15, 50]	.53 [.37, 1]	1 [.61, 1]	∞[a] [2.21, ∞]	0.47 [0, 0.63]	.82 [.66, .95]
Nonword	277.5 [268.5, 277.5]	1 [.53, 1]	.55 [.42, 1]	2.22 [1.73, ∞]	0[a] [0, 0.5]	.85 [.71, .95]
Gram. J.	0.88 [0.86, 0.92]	.82 [.68, 1]	.75 [.44, .95]	3.29 [1.75, 15.84]	0.24 [0, 0.4]	.82 [.67, .94]

Note. Optimal cutoff scores estimated from 37 participants (*n* = 36 for Word Classes). Derived measures of diagnostic accuracy are listed with 95% confidence intervals (CIs) representing percentiles 2.5 and 97.5 with 1,000 bootstrap iterations. PosLH = positive likelihood; NegLH = negative likelihood; AUC = area under the curve; Recall. = Recalling Sentences; ∞ = infinite; Word C. = word classes; EVIP = *Échelle de vocabulaire en images Peabody*; fLEX = fLEX verb production; F. Digit = forward digit span; Nonword = nonword repetition; Gram. J. = grammaticality judgment on subject–verb number agreement.

[a]For NegLH, 0 means perfect identification of participants with typical language, and for posLH, ∞ means perfect identification of participants with developmental language disorder.

Downloaded from: https://pubs.asha.org Université de Montreal on 01/07/2023, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

participants. The other subtasks had AUCs between .82 and .89, but because their AUCs dropped under .8 in their CIs, their accuracy only reached an acceptable classification. ROC plots for all seven subtasks are available in Supplemental Material S1, Figure S1.

Combination of Subtasks

We fitted a regularized logistic regression model with group classification (TL or DLD) as a dependent variable and our selection of seven subtask scores as independent variables. Since the variables were set to the same scale before model fit, model coefficients further from zero suggest a stronger contribution of the subtask. Results showed that Recalling Sentences contributed the most to the model (coefficient = −1.718, mean AUC gain = .48). The only other contributing subtask was fLEX verb production (coefficient = −0.510, mean AUC gain = .014). Since Recalling Sentences is our most discriminating subtask and alone provides almost perfect discrimination between our groups, it shrunk the contribution of the other subtasks to zero (see Supplemental Material S1, Section 2.3 and Table S3, for more detail on this analysis). To assess the contribution of the other subtasks, we therefore removed the most contributing subtask, Recalling Sentences, and ran a second model for the remaining six subtask scores. The resulting coefficients of this second model are shown in Table 5, along with relative variable importance. The coefficients for nonword repetition and grammaticality judgment shrunk to zero, indicating that these variables were eliminated by the regularization procedure. Of the remaining four subtask scores, fLEX verb production showed the largest contribution; including this score improved the AUC of the model by an average of .2, as opposed to including a randomly permuted vector containing the same values. The model improvement was smaller for Word Classes (mean AUC gain = .028) and EVIP (0.029) and almost null for Forward Digit Span (0.001). When used to classify the participants between TL and DLD groups, the final model had a sensitivity of .88, a specificity of 1, and an AUC of .98. Of 36 participants with no missing values, the model accurately classified 34 and produced two false negatives (two participants with DLD were classified as being the TL group).

Table 5. Coefficients and variable importance for the regularized logistic regression model.

Subtasks	Coefficient	Variable contribution (*SD*)
fLEX verb production	−1.00	0.200 (0.067)
EVIP	−0.424	0.029 (0.018)
Word Classes	−0.423	0.028 (0.020)
Forward Digit Span	−0.169	0.001 (0.010)
Nonword repetition	0	0 (0)
Grammaticality judgment (sub)irregular verbs	0	0 (0)

Note. Coefficients and variable importance produced with lambda parameter previously set by a leave-one-out cross-validation procedure. Variable contribution indicates the mean difference in the model's area under the curve when the variable is permuted; results shown are for 100 permutations. EVIP = *Échelle de vocabulaire en images Peabody*.

Discussion

This study first aimed to identify reliable tasks used in research and clinical settings that discriminate French-speaking adolescents with DLD from their peers with TL. Based on 20 subtasks administered to 37 (pre)teenagers with and without DLD, we found seven subtasks that displayed high levels of diagnostic accuracy. Three of them showed outstanding diagnostic accuracy: first was Recalling Sentences, followed by the receptive Word Classes task and the fLEX subject–verb number agreement production task. A second objective was to assess which linguistic domain(s) more specifically represented areas of weakness in our participants with DLD. To do so, we compared our most discriminating subtasks directly to see which one(s) contributed the most to identify French teenagers with DLD. We found that the fLEX subject–verb agreement production subtask assessing morphosyntactic skills contributed more to the model's diagnostic accuracy than subtasks assessing lexicosemantics and phonological working memory, revealing morphosyntax as a special area of weakness in French-speaking teenagers with DLD.

Discriminating Tasks for Québec French–Speaking Adolescents With DLD

Of the 20 subtasks administered to participants, seven subtasks were found to be informative about the group to which the participants belonged. The TL group performed better than the DLD group on all seven subtasks. The superiority of the TL group, younger on average than the DLD group, could partially be accounted for by the fact that their percentile scores were age-standardized for the EVIP, Recalling Sentences, Word Classes, and Forward Digit Span subtasks. However, the TL scores on the experimental tasks (fLEX subject–verb agreement production, nonword repetition, and grammaticality judgment) were also higher, even though they were not age-standardized. In the case of the subject–verb agreement production and grammaticality judgment tasks, both targeting subject–verb agreement, this suggests that this type of agreement is better mastered by (pre)adolescents with TL than by adolescent participants with DLD,

Downloaded from: https://pubs.asha.org Université de Montreal on 01/07/2023, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

for whom it is clearly impaired. This finding is in line with the studies of Rose and Royle (1999) and Franck et al. (2004). As these latter authors did, we found 5% of errors, on average, in the TL group (i.e., one error on 20 verbs) on subject–verb agreement production. However, looking at individual score distributions (see Figure 1), half of the participants with TL actually made more than just one error. Our results suggest that typically developing French-speaking children and adolescents have not yet fully acquired the production of subject–verb agreement. Our results thus suggest that small numbers of errors in subject–verb agreement elicitation tasks should not be considered indicators of DLD in children ages 8–14 years.

We calculated optimal cutoff scores that best classified our participants in our two groups for each of the seven discriminating tasks. Recommended cutoff scores for standardized subtasks typically start at −1 *SD* (or the 16th percentile) for mild language impairment. We found one score close to the recommended cutoff score for the CELF Forward Digit Span task (17th percentile). However, we found much higher cutoff scores for other standardized subtasks assessing linguistic skills, including the EVIP (59th percentile), Recalling Sentences (43.5th percentile), and Word Classes (62.5th percentile), with the 16th percentile also missing from the bootstrapped CIs. In a nutshell, all our participants exceeded expected performance, as our optimal cutoff scores were close to the average scores in published norms, that is, 50th percentile. What could explain these surprising results? A first interpretation could be that our sample was composed of particularly high-performing participants for their age. This is implausible because most of our adolescents with DLD had important language impairments as evidenced by their attendance at a school with special accommodations. A second explanation could be that the published norms were conducted on surprisingly low-performing groups, which is unlikely, at least for the CELF, given that it was based on a considerable number of children; 520 francophones from Québec aged 4–16 years. Another explanation for our high cutoff scores could be that the French CELF test, which three of our four standardized subtasks were taken from, has poor psychometric properties resulting in inadequate recommended cutoff scores; however, the English version of this test was identified as one of the recommended tests to evaluate language, based on its good psychometric qualities (Denman et al., 2017). The most compelling explanation for these results would be the problematic adaptation of language stimuli in these tests, which we believe did not target linguistic constructs important for assessing Québec French language development. Along with the studies of Godard and Labelle (1995) and Elin Thordardottir et al. (2011), our finding supports the possibility that French versions of English tests used by Québec SLPs underestimate language difficulties in teenagers with DLD. These results also support the argument that low scores on standardized tests should not be a criterion to assess disorder severity (Bishop et al., 2017; Breault et al., 2019; Tessier et al., 2017). Future adaptions of English standardized language batteries should take into account linguistic features that are known to be difficult for children with DLD, for instance, subject–verb agreement production, as suggested by our results, or past tense production in regular and irregular verbs (Royle et al., 2018).

Based on our optimal cutoff scores, we calculated diagnostic accuracy statistics. Of the seven subtasks we selected, three were found to have outstanding discriminating ability, with AUCs above .90, and maintaining excellent AUC values in their CI. These subtasks also had good-to-fair sensitivity and specificity, as well as strong, moderate, or near-moderate effects on both positive and negative likelihood ratios. Unsurprisingly, Recalling Sentences discriminated best between our teenage participants with DLD versus TL, with an AUC of .98. This task had been proven to be a powerful diagnostic tool with French children aged 5 years (Elin Thordardottir et al., 2011) and 7–12 years (Leclercq et al., 2014), and we now can confirm that it is still highly relevant when assessing DLD in 14-year-olds. One explanation for this could be that sentence repetition tasks cover two differentiated language dimensions, that is, (morpho)syntax and lexicosemantics (Lonigan & Milburn, 2017; Tomblin & Zhang, 2006). This type of task has been shown to assess these two linguistic areas almost equally (Leclercq et al., 2014). Future studies could investigate whether any of these two linguistic domains play a greater role in the assessment of DLD in children and adolescents.

Our second-best diagnostic task was Word Classes (AUC = .93). This result is consistent with studies by McGregor et al. (2013) and Rice and Hoffman (2015), which also found lexicosemantic deficits in teens with DLD, and suggests that future research should consider including lexical–semantic skills as a marker for DLD. Furthermore, this result highlights that when assessing lexicosemantic skills in teenagers, the use of multiword tasks such as Word Classes is more accurate in diagnosing DLD compared with a single-word task such as the EVIP. Indeed, even if the EVIP's AUC (.89; see Table 4) was close to that of Word Classes, the EVIP's AUC was below .80 when looking at the CI, which is considered only acceptable, whereas the Word Classes' CI AUC remained excellent and over .80. The third subtask with the best AUC was fLEX subject–verb agreement production. This finding suggests that subject–verb agreement production deficits are impaired in French teenagers with DLD. Recall that, in the fLEX test, the same verbs were used in the expressive vocabulary task and in the subject–verb agreement production task. No difference was found between groups in the vocabulary task, implying that verb

Downloaded from: https://pubs.asha.org Université de Montreal on 01/07/2023, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

lexical representation was similarly accessible in both groups. Our results suggest that participants with DLD had difficulty inflecting the verbs they could nevertheless name in the infinitive, pointing toward morphosyntactic deficits. Elin Thordardottir et al. (2011) found that receptive morphosyntactic assessment, but not production, was one of the five best tasks that accurately discriminated French-speaking 5-year-olds with and without DLD. This difference might be because, as they age, children with DLD better master comprehension, but maintain difficulties in verb production. This finding is also coherent with Paradis and Crago's (2001) "extended optional default" account. Although some of our participants with TL still made one or two errors in this subtask, the majority of the DLD group made several. In both groups, the singular was produced rather than the plural form.

The two tasks assessing phonological working memory, Forward Digit Span and nonword repetition, had lower diagnostic accuracy. Even if their AUCs were above .80, which is considered excellent, their CIs were only acceptable. A French study that assessed phonological working memory with number or nonword repetition has found lower performance in teens with DLD when compared with their peers with TL, which is consistent with our results (Arslan et al., 2020). However, the authors did not analyze diagnostic accuracy as we did. It may come as a surprise that the two tasks assessing phonological working memory have not demonstrated great discrimination between groups. Deficits in this area have been identified as a feature of DLD across languages (Leonard, 2014) and are expected in DLD based on the procedural deficit hypothesis, as phonological working memory is supported by the procedural memory system, which is hypothesized to be impaired in DLD (Ullman et al., 2020; Ullman & Pierpont, 2005). The best explanation for our results would be our choice of task to assess phonological working memory. The nonword repetition task we used was probably not appropriate for the age range we studied because the task was too easy, having being designed for kindergarten children. Indeed, both groups performed almost at ceiling. A more age-appropriate task, for example, a task with more complex syllable structures that evaluates not only phonological working memory but phonological skills in general (Tuller et al., 2018), could have achieved better diagnostic accuracy.

Of typical tasks used in research on DLD, only one discriminated between our groups, namely, grammaticality judgments on verbs carrying audible subject–verb agreement. This finding shows that this kind of morphosyntactic task taps into deficits in teenagers with DLD and is consistent with many studies that found similar results (Haebig et al., 2017; Maillart & Schelstraete, 2005; Miller et al., 2008; Noonan et al., 2014; Poulin et al., 2015; Rose & Royle, 1999). However, these studies did not evaluate tasks' diagnostic accuracy. Our results show that this morphosyntactic judgment task's diagnostic accuracy was low, as illustrated by the AUC that was only acceptable when looking at the CI, but recall that our task was presented in the context of an ERP study (Courteau et al., under revision) and not in a typical clinical setting. Given the extensive evidence of impaired grammaticality judgment skills in DLD, future studies should focus on assessing the diagnostic accuracy of these types of tasks in clinical settings.

Linguistic Domains Identified as Areas of Weakness in French Teenagers With DLD

Using a multivariable model, we were able to explore the relative contribution of each subtask to the model's total AUC, which was .98, and corresponded to an outstanding classification. Only three subtasks were considered to contribute significantly to it. The subtasks assessing lexicosemantics—Word Classes and EVIP—contributed respectively .028 and .029 AUC to the model, whereas the fLEX subject–verb agreement production contributed .20 AUC. These results strongly imply that an impairment in morphosyntax is a more discriminant characteristic of French-speaking adolescents with DLD, as opposed to impairments in lexical semantics. Our results are coherent with that of Leonard (2014), who identified grammatical computation as an area of weakness in DLD across languages, but not lexicosemantics. However, Leonard suggests that children speaking Romance languages (which include French) do not show the severe deficits in number agreement that are found in English-speaking children. This is coherent with Hamann et al.'s (2003) study, which found that 7-year-old French-speaking children with DLD made very few nonfinite form errors in their spontaneous speech. This is also partly in line with Moscati et al.'s (2020) fragile computation of agreement hypothesis, which states that subject–verb agreement difficulties in DLD increase as a function of the complexity of the agreement configuration, based on Italian-speaking children with DLD. Nevertheless, our results suggest that subject–verb agreement errors are observable in French-speaking teenagers with DLD in simple agreement configurations, as the one used in our verb production task. Furthermore, a study published by Leonard's team (Leonard et al., 1992) found disproportionate difficulties in plural subject–verb agreement production (and possibly comprehension)[8] for Italian-speaking children with DLD aged 4;9–5:11. An interpretation of the lack of consistent effects of number agreement as a reliable marker could be that, in French

[8]Statistical comparisons were only made with the younger MLU-matched group. The age-matched group was near perfect in their results.

Downloaded from: https://pubs.asha.org Université de Montreal on 01/07/2023, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

and some other languages, it is only a reliable marker in teenagers due to idiosyncratic properties of the morphological system. Furthermore, the different task types used in these studies could explain the discrepancies between them, that is, a comparison of spontaneous versus elicited production (Prigent et al., 2015) suggests that teens with DLD avoid using complex morphosyntactic structures in their spontaneous speech. Considering this, future studies in the Romance languages should focus on older populations and use elicited production tasks when investigating subject–verb number agreement in DLD.

Weaknesses in (morpho)syntax with concurrent relative strength in lexicosemantics are consistent with the procedural deficit hypothesis (Ullman et al., 2020; Ullman & Pierpont, 2005); as in persons with DLD, the former is supported by an impaired procedural memory system, and the latter is supported by a well-developed declarative memory system. Note that this hypothesis does not predict that irregular and subregular verbs, such as those used in the fLEX subject–verb agreement production subtask, should be particularly impaired in DLD as it assumes that these verbs can be easily processed by the declarative system because they can be chunked (i.e., stored) at the word level. However, at the sentence level, this subtask involves processing agreement, an abstract morphosyntactic feature, and relies on the procedural system, which is expected to be impaired (Steinhauer & Ullman, 2002).

This finding is reminiscent of what was found in English-speaking preschoolers (e.g., Leonard et al., 1999). However, Elin Thordardottir and Namazi (2007) did not find morphosyntactic difficulties to be salient when compared with lexicosemantic skills in spontaneous speech of 5-year-old French-speaking children with DLD. In our view, their finding is due to the fact that (a) morphosyntactic skills are still developing in preschool children, at least when focusing on subject–verb agreement, and (b) spontaneous utterances did not provide contexts in which they could target deficits specific to DLD (see Royle & Reising, 2019, for a comparison of spontaneous and elicited speech). Our results carry implications for publishers of French tests targeting adolescents, who should consider including a subject–verb number agreement production task focusing on subregular and irregular verbs.

Together with Elin Thordardottir et al. (2011), our findings offer clear evidence that children with DLD's linguistic skills change and evolve with age. Indeed, these authors identified two of our same subtasks, EVIP and nonword repetition, as the combination that provided the best diagnostic accuracy for 5-year-olds with an AUC of .98, whereas these tasks had little or no contribution to our adolescent data model.

Based on subject–verb agreement skills, our study demonstrated that morphosyntactic impairments are a reliable marker of French DLD in teenagers and is a special area of weakness when compared with lexicosemantics, similar to what has been found in English for younger individuals with DLD. In order to investigate the scope of morphosyntax as a marker of DLD in French, future studies should compare morphosyntactic skills beyond number agreement and ensure that the psychometric properties of their tests are valid.

Study Limitations

There are four main potential limitations to this study. A first limitation concerns our sample size. Considering our small sample, outliers could have caused accidental characteristics to be confused with actual trends. We used a robust statistical methodology to mitigate this problem. Participants in the control group were younger, on average, than those in the DLD group and also covered a broader age range. Despite this, we found seven subtasks where their performance was superior to that of teenagers with DLD. However, if we had had a control group of the same age as those in the DLD group, perhaps additional tasks in our selection would have discriminated groups more robustly. Furthermore, our sample included only French monolinguals. Bilingualism in Québec has been evaluated to be around 45% and growing (Statistics Canada, 2017). Identification of reliable tasks to assess bilingual French teenagers with DLD would be relevant for the province of Québec. A fourth potential issue concerns the selection of tests and tasks used in this study, which were not always adapted to the older age of our participants. Indeed, both groups performed almost at ceiling on many experimental tasks, partly explaining why only seven of 20 subtasks were relevant. To avoid this, we could have selected additional tasks that directly assess promising markers for older French-speaking children with DLD. For instance, within morphosyntax, complex syntax skills have been shown to be impaired in children ages 5;0–14;6 (Delage & Frauenfelder, 2020), as has the use of clitic markers for 6- to 13-year-olds (Stanford et al., 2019). Since the ability to use complex syntax and clitics was assessed as part of the Recalling Sentences task, we did not assess their diagnostic accuracy as subtasks in our analyses. Future studies should focus on evaluating the clinical potential of these promising markers.

Conclusions

Our study contrasted two types of language assessment tasks, namely, clinical and research tools, that, to our knowledge, have not yet been directly compared in teenagers. This research can be considered as a first step toward identifying psycholinguistic markers of French-

Downloaded from: https://pubs.asha.org Université de Montreal on 01/07/2023, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

speaking adolescents with DLD. Taken together, our findings indicate that French-speaking adolescents with DLD still have deficits in oral language as basic as subject–verb agreement production. These deficits should be addressed in SLP intervention and, to a greater extent, in regular classroom settings. Although the focus in high school education is generally on written language, it is essential that intervention for adolescents with DLD continue to target oral language, as this remains the source of their difficulties whether in the oral or written form.

Data Availability Statement

The data sets analyzed during this study are available from the corresponding author on reasonable request.

Acknowledgments

This project was funded by SSHRC Insight grants to Royle, Steinhauer, et al. (435-2015-1280) and to Steinhauer and Royle (435-2013-0583), by an FQRSC grant to White, Steinhauer, Royle, et al. (2016-SE-188196) and by graduate PhD scholarships to Courteau (CRBLM, SSHRC, and FRQSC). The authors gratefully acknowledge the schools and families who participated in our study for their time and involvement in our project. They would like to thank the reviewers for taking the time and effort to review this article. They are sincerely appreciative of all the valuable comments and suggestions that have helped them improve the quality of this article.

References

Abdi, H. (2010). Holm's sequential Bonferroni procedure. *Encyclopedia of Research Design, 1*(8), 1–8.

Altmann, A., Toloşi, L., Sander, O., & Lengauer, T. (2010). Permutation importance: A corrected feature importance measure. *Bioinformatics, 26*(10), 1340–1347. https://doi.org/10.1093/bioinformatics/btq134

American Psychiatric Association. (2013). Neurodevelopmental disorders. In *Diagnostic and statistical manual of mental disorders* (5th ed., pp. 31–86). American Psychiatric Publishing. https://doi.org/10.1176/appi.books.9781585624836.jb01

Arslan, S., Broc, L., Olive, T., & Mathy, F. (2020). Reduced deficits observed in children and adolescents with developmental language disorder using proper nonverbalizable span tasks. *Research in Developmental Disabilities, 96,* 103522. https://doi.org/10.1016/j.ridd.2019.103522

Azhagusundari, B., & Thanamani, A. S. (2013). Feature selection based on information gain. *International Journal of Innovative Technology and Exploring Engineering, 2*(2), 18–21.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Balilah, A., Rafat, Y., & Archibald, L. (2019). Domain-specific and domain-general processing accounts in children with specific language impairment (SLI): Contribution of cross-linguistic evidence. In S. Hidri (Ed.), *English language teaching research in the Middle East and North Africa: Multiple perspectives* (pp. 383–407). Springer. https://doi.org/10.1007/978-3-319-98533-6_18

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

Bouchard, M. G., Fitzpatrick, E. M., & Olds, J. (2009). Psychometric analysis of assessment tools used with francophone children. *Canadian Journal of Speech-Language Pathology & Audiology, 33*(3), 129–139.

Bourget, M.-J. (1987). Variation phonétique dans l'emploi des pronoms de troisième personne en français [Phonetic variation in the use of third person pronouns in French]. In J. Auger, R. Grenier, R. Lapalme, J.-F. Montreuil, & P. Whitmore (Eds.), *Tendances actuelles de la recherche Sur la langue parlee* [Current trends in research on oral language] (Vol. B-166, pp. 129–140). Université Laval.

Breault, C., Béliveau, M.-J., Labelle, F., Valade, F., & Trudeau, N. (2019). Le trouble développemental du langage (TDL): Mise à jour interdisciplinaire [Developmental language disorder (DLD): An interdisciplinary update]. *Neuropsychologie clinique et appliquée, 3*(automne 2019), 64–81. https://doi.org/10.46278/j.ncacn.20190717

Brunner, E., & Munzel, U. (2000). The nonparametric Behrens–Fisher problem: Asymptotic theory and a small-sample approximation. *Biometrical Journal, 42*(1), 17–25. https://doi.org/10.1002/(SICI)1521-4036(200001)42:1<17::AID-BIMJ17>3.0.CO;2-U

Conti-Ramsden, G. (2008). Heterogeneity of specific language impairment in adolescent outcomes. In C. F. Norbury, J. B. Tomblin, & D. V. M. Bishop (Eds.), *Understanding developmental language disorders: From theory to practice* (pp. 115–129). Psychology Press.

Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *The Journal of Child Psychology and Psychiatry, 42*(6), 741–748. https://doi.org/10.1111/1469-7610.00770

Corsi, P. M. (1972). *Human memory and the medial temporal region of the brain* [Unpublished doctoral thesis]. McGill University.

Courteau, É., Martignetti, L., Royle, P., & Steinhauer, K. (2019). Eliciting ERP components for morphosyntactic agreement mismatches in perfectly grammatical sentences. *Frontiers in Psychology, 10,* 1152. https://doi.org/10.3389/fpsyg.2019.01152

Courteau, É., Royle, P., Gascon, A., Marquis, A., Drury, J. E., & Steinhauer, K. (2013). Gender concord and semantics processing in French children: An auditory ERP study. In S. Baiz, N. Goldman, & R. Hawkes (Eds.), *Proceedings of the 37th Annual BUCLD* (Vol. 1, pp. 87–99).

Courteau, É., Royle, P., & Steinhauer, K. (under revision). *Testing neurocognitive models of language processing in developmental language disorder*.

Daniel, T. A., Katz, J. S., & Robinson, J. L. (2016). Delayed match-to-sample in working memory: A BrainMap meta-analysis. *Biological Psychology, 120,* 10–20. https://doi.org/10.1016/j.biopsycho.2016.07.015

Delage, H., & Frauenfelder, U. H. (2020). Relationship between working memory and complex syntax in children with

Downloaded from: https://pubs.asha.org Université de Montreal on 01/07/2023, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

developmental language disorder. *Journal of Child Language, 47*(3), 600–632. https://doi.org/10.1017/S0305000919000722

Denman, D., Speyer, R., Munro, N., Pearce, W. M., Chen, Y.-W., & Cordier, R. (2017). Psychometric properties of language assessments for children aged 4–12 years: A systematic review. *Frontiers in Psychology, 8,* 1515. https://doi.org/10.3389/fpsyg.2017.01515

Dunn, L. M., Thériault-Whalen, C. M., & Dunn, L. M. (1993). *Échelle de vocabulaire en images Peabody: Adaptation française du Peabody Picture Vocabulary Test* [Peabody Picture Vocabulary Scale: French adaptation of the Peabody Picture Vocabulary Test]. PsyCan.

Duquette, A.-S., Courteau, E., & Royle, P. (2020). *Fidélité interjuge d'une étude sur le trouble développemental du langage et de la compréhension du langage oral* [Inter-rater fidelity of a study of developmental language disorder and oral language comprehension]. 53e Congrès Premier des stagiaires de recherche du 1er cycle de la Faculté de médecine, Montréal, QC. https://doi.org/10.13140/RG.2.2.27265.58722

Elin Thordardottir, Kehayia, E., Mazer, B., Lessard, N., Majnemer, A., Sutton, A., Trudeau, N., & Chilingaryan, G. (2011). Sensitivity and specificity of French language and processing measures for the identification of primary language impairment at age 5. *Journal of Speech, Language, and Hearing Research, 54*(2), 580–597. https://doi.org/10.1044/1092-4388(2010/09-0196)

Elin Thordardottir, & Namazi, M. (2007). Specific language impairment in French-speaking children: Beyond grammatical morphology. *Journal of Speech, Language, and Hearing Research, 50*(3), 698–715. https://doi.org/10.1044/1092-4388(2007/049)

Franck, J., Cronel-Ohayon, S., Chillier, L., Frauenfelder, U. H., Hamann, C., Rizzi, L., & Zesiger, P. (2004). Normal and pathological development of subject–verb agreement in speech production: A study on French children. *Journal of Neurolinguistics, 17*(2-3), 147–180. https://doi.org/10.1016/S0911-6044(03)00057-5

Friedmann, N., & Novogrodsky, R. (2009). Subtypes of SLI: SySLI, PhoSLI, LeSLI, and PraSLI. In A. Gavarró & M. J. Freitas (Eds.), *Language acquisition and development: Proceedings of GALA 2007* (pp. 205–217). Cambridge Scholars.

Godard, L., & Labelle, M. (1995). Utilisation de l'ÉVIP avec Une population Québécoise [Use of the EVIP with a Québec population]. *Fréquences, 7,* 18–21.

Greenwell, B. M., & Boehmke, B. C. (2020). Variable importance plots—An introduction to the VIP package. *The R Journal, 12*(1), 343. https://doi.org/10.32614/RJ-2020-013

Haebig, E., Weber, C., Leonard, L. B., Deevy, P., & Tomblin, J. B. (2017). Neural patterns elicited by sentence processing uniquely characterize typical development, SLI recovery, and SLI persistence. *Journal of Neurodevelopmental Disorders, 9*(1), 22. https://doi.org/10.1186/s11689-017-9201-1

Hajian-Tilaki, K. (2013). Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian Journal of Internal Medicine, 4*(2), 627–635.

Hamann, C., Ohayon, S., Dubé, S., Frauenfelder, U. H., Rizzi, L., Starke, M., & Zesiger, P. (2003). Aspects of grammatical development in young French children with SLI. *Developmental Science, 6*(2), 151–158. https://doi.org/10.1111/1467-7687.00265

Hastie, T., Qian, J., & Tay, K. (2016). *An introduction to glmnet*. https://glmnet.stanford.edu/articles/glmnet.html

Hui, W., Gel, Y. R., Gastwirth, J. L., & Miao, W. (2020). *Brunnermunzel.*

Jaeschke, R., Guyatt, G. H., Sackett, D. L., Guyatt, G., Bass, E., Brill-Edwards, P., Browman, G., Cook, D., Farkouh, M., Gerstein, H., Haynes, B., Hayward, R., Holbrook, A., Juniper, E., Lee, H., Levine, M., Moyer, V., Nishikawa, J., Oxman, A., . . . Wilson, M. (1994). Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? *JAMA, 271*(9), 703–707. https://doi.org/10.1001/jama.1994.03510330081039

Krippendorff, K. (2004). Reliability in content analysis. *Human Communication Research, 30*(3), 411–433. https://doi.org/10.1111/j.1468-2958.2004.tb00738.x

Kuhn, M. (2009). The caret package. *Journal of Statistical Software, 28*(5).

Leclercq, A.-L., Quémart, P., Magis, D., & Maillart, C. (2014). The sentence repetition task: A powerful diagnostic tool for French children with specific language impairment. *Research in Developmental Disabilities, 35*(12), 3423–3430. https://doi.org/10.1016/j.ridd.2014.08.026

Leonard, L. B. (2014). Specific language impairment across languages. *Child Development Perspectives, 8*(1), 1–5. https://doi.org/10.1111/cdep.12053

Leonard, L. B., Caselli, M. C., Bortolini, U., McGregor, K. K., & Sabbadini, L. (1992). Morphological deficits in children with specific language impairment: The status of features in the underlying grammar. *Language Acquisition, 2*(2), 151–179. https://doi.org/10.1207/s15327817la0202_2

Leonard, L. B., Miller, C., & Gerber, E. (1999). Grammatical morphology and the lexicon in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 42*(3), 678–689. https://doi.org/10.1044/jslhr.4203.678

Lonigan, C. J., & Milburn, T. F. (2017). Identifying the dimensionality of oral language skills of children with typical development in preschool through fifth grade. *Journal of Speech, Language, and Hearing Research, 60*(8), 2185–2198. https://doi.org/10.1044/2017_JSLHR-L-15-0402

Maillart, C., & Schelstraete, M.-A. (2005). Grammaticality judgment in French-speaking children with specific language impairment. *Journal of Multilingual Communication Disorders, 3*(2), 103–109. https://doi.org/10.1080/14769670500066479

McGregor, K. K., Oleson, J., Bahnsen, A., & Duff, D. (2013). Children with developmental language impairment have vocabulary deficits characterized by limited breadth and depth. *International Journal of Language & Communication Disorders, 48*(3), 307–319. https://doi.org/10.1111/1460-6984.12008

Miller, C. A., Leonard, L. B., & Finneran, D. (2008). Grammaticality judgements in adolescents with and without language impairment. *International Journal of Language & Communication Disorders, 43*(3), 346–360. https://doi.org/10.1080/13682820701546813

Monetta, L., Desmarais, C., MacLeod, A. A., St-Pierre, M.-C., Bourgeois-Marcotte, J., & Perron, M. (2016). Recension des outils franco-québécois pour l'évaluation des troubles du langage et de la parole [Review of Quebec-French tools for the assessment of speech and language disorders]. *Revue canadienne d'orthophonie et d'audiologie, 40*(2), 165–175.

Moscati, V., Rizzi, L., Vottari, I., Chilosi, A. M., Salvadorini, R., & Guasti, M. T. (2020). Morphosyntactic weaknesses in developmental language disorder: The role of structure and agreement configurations. *Journal of Child Language, 47*(5), 909–944. https://doi.org/10.1017/S0305000919000709

Noonan, N. B., Redmond, S. M., & Archibald, L. M. D. (2014). Contributions of children's linguistic and working memory proficiencies to their judgments of grammaticality. *Journal of*

Downloaded from: https://pubs.asha.org Université de Montreal on 01/07/2023, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Speech, Language, and Hearing Research, 57(3), 979–989. https://doi.org/10.1044/2014_JSLHR-L-12-0225

Paradis, J., & Crago, M. (2001). The morphosyntax of specific language impairment in French: An extended optional default account. *Language Acquisition, 9*(4), 269–300. https://doi.org/10.1207/S15327817LA0904_01

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

Poulin, M.-J., Marquis, A., & Royle, P. (2015). Étude de faisabilité portant Sur l'évaluation de la production et de la compréhension du langage oral en français [Feasibility study on the assessment of oral language production and comprehension in French]. In M. Pomerleau & E. M. Gendron-Pontbrian (Eds.), *ScriptUM: La revue du colloque VocUM* (Vol. 1, pp. 54–68).

Pourquié, M. (2017). Evaluation du lexique et de la flexion verbale dans le trouble primaire du langage [Assessment of lexicon and verbal inflection in primary language impairment]. *ScriptUM: La revue du colloque VocUM, 3,* 35–53. https://scriptum.vocum.ca/index.php/scriptum/article/view/47

Pourquié, M., Courteau, É., Duquette, A.-S., & Royle, P. (under revision). *Verb inflection and argument structure processing in French adolescents with DLD* [Manuscript submitted for publication].

Prigent, G., Parisse, C., Leclercq, A.-L., & Maillart, C. (2015). Complexity markers in morphosyntactic productions in French-speaking children with specific language impairment (SLI). *Clinical Linguistics & Phonetics, 29*(8–10), 701–718. https://doi.org/10.3109/02699206.2015.1020451

Quémart, P., & Maillart, C. (2016). The sensitivity of children with SLI to phonotactic probabilities during lexical access. *Journal of Communication Disorders, 61,* 48–59. https://doi.org/10.1016/j.jcomdis.2016.03.005

Redmond, S. M., Ash, A. C., & Hogan, T. P. (2015). Consequences of co-occurring attention-deficit/hyperactivity disorder on children's language impairments. *Language, Speech, and Hearing Services in Schools, 46*(2), 68–80. https://doi.org/10.1044/2014_LSHSS-14-0045

Rice, M. L., & Hoffman, L. (2015). Predicting vocabulary growth in children with and without specific language impairment: A longitudinal study from 2;6 to 21 years of age. *Journal of Speech, Language, and Hearing Research, 58*(2), 345–359. https://doi.org/10.1044/2015_JSLHR-L-14-0150

Rietveld, T., & van Hout, R. (2015). The t test and beyond: Recommendations for testing the central tendencies of two independent samples in research on speech, language and hearing pathology. *Journal of Communication Disorders, 58,* 158–168. https://doi.org/10.1016/j.jcomdis.2015.08.002

Rose, Y., & Royle, P. (1999). Uninflected structure in familial language impairment: Evidence from French. *Folia Phoniatrica et Logopaedica, 51*(1–2), 70–90. https://doi.org/10.1159/000021482

Royle, P., & Reising, L. (2019). Elicited and spontaneous determiner phrase production in French speaking children with developmental language disorder. *Canadian Journal of Speech Language Pathology and Audiology, 43*(3), 167–187.

Royle, P., St-Denis, A., Mazzocca, P., & Marquis, A. (2018). Insensitivity to verb conjugation patterns in French children with SLI. *Clinical Linguistics & Phonetics, 32*(2), 128–147. https://doi.org/10.1080/02699206.2017.1328706

RStudio Team. (2020). *RStudio: Integrated development for R*. http://www.rstudio.com/

Secord, W. A., Wiig, E., Boulianne, L., Semel, E., & Labelle, M. (2009). *Évaluation clinique des notions langagières fondamentales– Version pour francophones du Canada (CELF CDN-F)* [Clinical Assessment of Language Fundamentals—French Canadian Version (CELF CDN-F)]. The Psychological Corporation.

Stanford, E., Durrleman, S., & Delage, H. (2019). The effect of working memory training on a clinical marker of French-speaking children with developmental language disorder. *American Journal of Speech-Language Pathology, 28*(4), 1388–1410. https://doi.org/10.1044/2019_AJSLP-18-0238

Statistics Canada. (2016). *Mother tongue for the total population excluding institutional residents (100% data)*. Census Profile Québec and Canada. https://www12.statcan.gc.ca/

Statistics Canada. (2017). *English–French bilingualism reaches new heights*. https://www12.statcan.gc.ca/census-recensement/2016/as-sa/98-200-x/2016009/98-200-x2016009-eng.cfm

Steinhauer, K., & Ullman, M. T. (2002). Consecutive ERP effects of morpho-phonology and morpho-syntax. *Brain and Language, 83,* 62–65.

Tessier, A., Valade, S., Beaudoin, I., Nadeau, C. L., & Normandin, É. (2017). *Organisation du continuum et de la dispensation des services aux enfants âgés de 2 à 9 ans présentant un trouble développemental du langage (trouble primaire du langage)* [Organization of the continuum and delivery of services to children aged 2 to 9 years with a developmental language disorder (primary language impairment)] (p. 107). Gouvernement du Quebec. http://collections.banq.qc.ca/ark:/52327/3185271

Thiele, C., & Hirschfeld, G. (2020). *Cutpointr: Improved estimation and validation of optimal cutpoints in R*. ArXiv Preprint ArXiv:2002.09209.

Tomblin, J. B., & Zhang, X. (2006). The dimensionality of language ability in school-age children. *Journal of Speech, Language, and Hearing Research, 49*(6), 1193–1208. https://doi.org/10.1044/1092-4388(2006/086)

Tuller, L., Hamann, C., Chilla, S., Ferré, S., Morin, E., Prevost, P., dos Santos, C., Abed Ibrahim, L., & Zebib, R. (2018). Identifying language impairment in bilingual children in France and in Germany. *International Journal of Language & Communication Disorders, 53*(4), 888–904. https://doi.org/10.1111/1460-6984.12397

Tuller, L., Henry, C., Sizaret, E., & Barthez, M.-A. (2012). Specific language impairment at adolescence: Avoiding complexity. *Applied Psycholinguistics, 33*(1), 161–184. https://doi.org/10.1017/S0142716411000312

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

Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex, 41*(3), 399–433. https://doi.org/10.1016/S0010-9452(08)70276-4

Zhang, J., & Mueller, S. T. (2005). A note on ROC analysis and non-parametric estimate of sensitivity. *Psychometrika, 70*(1), 203–212. https://doi.org/10.1007/s11336-003-1119-8

Downloaded from: https://pubs.asha.org Université de Montreal on 01/07/2023, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions