Supplemental Material S1.

1. Detailed task presentation

The tasks used in this study can be classified into two categories and will be briefly described in the following order: 1) those commonly used by speech-language pathologists (SLP) in clinical settings to assess language skills of Québec French adolescents, and 2) those used in research on developmental language disorder (DLD).

Three tasks were selected from the CELF-IV$^{\text{cnd-F}}$ French version standardized among Québec-French speakers ages 4 to 16 (Secord et al., 2009) and were administered as recommended by the manual. The Recalling Sentences task, where participants needed to repeat orally presented sentences without any word changes, assessed lexicosemantics, morphosyntax and phonological working memory skills (Leclercq et al., 2014). The Word Classes task assessed ability to understand the lexicosemantic relationships between orally presented words by choosing two words that go together in a choice of four (receptive subtask) and to explain this relationship (expressive subtask). The number repetition tasks consisted in a forward and backward digit span. We also used the non-standardized French Québec Nonword repetition task (developed by Courcy, Elin Thordardottir et al., 2011) which consists of 40 words ranging in length from two to five syllables. Scoring followed the task’s recommendation: phoneme omissions and substitutions were counted as incorrect, while distortions and additions did not result in point loss. A point was given for each repeated phoneme, with a maximum of 280, and we used the total repeated phonemes as participants’ score. Neurotypical adults repeat between 277 and 279 phonemes correctly, as revealed by a task pre-validation with 10 French-speakers having French as their daily language (Duquette et al., 2020). We chose the EVIP task, a standardized Canadian French version of the Peabody Vocabulary Test for 2:5 to 18 years old (Dunn et al., 1993) to evaluate the receptive vocabulary. In this, participants choose among four pictures the one matching a word spoken by the experimenter. Expressive vocabulary was assessed with an action (verb) naming task taken from the French version of the fLEX test (task 2, see Pourquié et al., 2017), where participants had to describe with a verb each of 30 actions depicted on pictures. Subject-verb number agreement production and comprehension skills were assessed through tasks three and four of the fLEX test (ibid), which each contained 35 items. To target verbs that had an audible agreement number cue[1] on the verb's ending (e.g., *il rugit* [ilʁyʒi] “he roars” vs. *ils rugissent* [ilʁyʒɪs] “they roar”), we rated a subset of 20 items from the original task, bringing the maximum score to 20. The expressive task assessed sentence production of inflected verbs in the present tense, either in the singular or plural depending on the number of agents depicted on the picture. The receptive task assessed understanding of inflected verbs in the singular and plural: the participant chose among four pictures the one that matched a sentence spoken by the experimenter. This sentence-picture matching paradigm contained one target image and three foils: a) a number agreement-error, e.g., one lion roaring for *they roar*, b) a lexical error on the verb, e.g., one lion sleeping, or c) a combined lexical and number foil, e.g., two lions sleeping. We used two grammaticality judgment tasks where an alien comes to Québec to learn French and sometimes makes mistakes (Courteau et al., 2013). These data were taken from an off-line grammaticality judgment task and an event-related potentials (ERP) session (Courteau et al., under revision). During the off-line task, participants had to listen to pre-recorded sentences while looking at pictures and judge if sentences correctly described the picture or not by answering yes or no. The first task was adapted from Poulin et al. (2015). Participants listened to 16 sentences while watching pictures that were either correct (4) or contained errors targeting the noun phrase (*n* = 12). Errors included auditory-visual lexicosemantic mismatches on nouns (e.g., visual [BROWN SHOE ON TABLE], *Je vois un !train brun...*, “I see a brown !train…,” *n* = 4) and morphosyntactic gender-agreement errors on the determiner (e.g., *Je vois *<u>la</u>*

[1] Verbs were either sub-regular or irregular verbs from the 2nd and 3rd conjugation groups.

soulier vert …, "I see *the.F shoe.M green.M…," *n* = 4) or the adjective (e.g., *Je vois le soulier *<u>verte</u> …,* "I see the.M shoe.M *green.F…," *n* = 4). In the second task run during the EEG recording, participants listened to 300 sentences while watching pictures that were either a match (150) or contained errors that targeted the verb (150) and judged if the visuo-auditory pairs were a match or not using a button press. Lexicosemantic errors were created with a verb that did not match the depicted action (*n* = 30, e.g., visual [A WOMAN SINGS], *… elle !nage dans la piscine publique,* "she !swims in the public pool"). Subject-verb number agreement errors were created by varying the number of visually presented agents and morphosyntactic number cues in the auditory stimuli. All auditory cues were perfectly grammatical (Courteau et al, 2019). This was operationalized using either verbs with regular agreement morphophonology, where the plural number cue "s" [z] is created by the liaison between the pronoun plural form and verb's vowel onset (*n* = 60, e.g., visual [A GIRL EATS], *Au dessert, *elles‿aiment [ɛlzɛm] la mousse au chocolat,* "For dessert, *they like chocolate mousse") or with irregular and subregular verbs whose number cue was audible on the verb ending (*n* = 60, e.g., visual [A LION ROARS], *En soirée, ils *rugissent* [ilʁyʒɪs] *dans la savane,* "In the evening, *they roar in the jungle"). Further, we tested participants with commonly used tasks in DLD research assessing nonverbal visual working memory.[2] We used 4 computer-based nonverbal working memory tasks (Cognitive Experiments IV v2 pack of the Presentation® software, Version 18.0, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com). Of these was the forward and backward Corsi Blocks tasks (Corsi, 1972), where a sequence of highlighted squares is presented on the computer screen, and the participant must recreate the sequence using the mouse in forward or backward order. We used a delayed match-to-sample task of non-verbal stimuli (Daniel et al., 2016) where a form made of sixteen squares is displayed, and after a delay of 1 or 5 seconds, the participant must recall the form by choosing the right one in a choice of two. See Table S1 for the type of score available (raw, aged-based percentile, *A*-score), and the underlying linguistic or working memory subdomains assessed by each subtask, based on the tests' manual when available, or the literature.

Table S1. *Complete list of subtasks and the underlying linguistic and cognitive subdomain they assessed.*

Subtasks	Scores	Linguistic and cognitive subdomains
Subtasks used in clinical settings		
Recalling	Raw, Pcl	Lexicosemantics, Morphology and Syntax (Leclercq et al., 2014)
Word Classes	Raw, Pcl	Lexicosemantic classes' relationship (CELF-IV$^{cnd\text{-}F}$)
Forward Digit	Raw, Pcl	Verbal working memory (CELF-IV$^{cnd\text{-}F}$; Baddeley et al., 2000)
Backward Digit	Raw, Pcl	Verbal working memory (CELF-IV$^{cnd\text{-}F}$; Baddeley et al., 2000)
Nonword Rep.	Raw	Verbal working memory (Gathercole et al., 1994)
EVIP	Raw, Pcl	Lexicosemantics: receptive vocabulary (Dunn et al., 1993)
Action naming	Raw	Lexicosemantics: lexical access of verbs (Pourquié et al., 2017)
S-V verb prod.	Raw	Morphosyntactic processing of verbs (Pourquié et al., 2017)
S-V verb comp.	Raw	Morphosyntactic processing of verbs (Pourquié et al., 2017)
Subtasks used in research on DLD		
J. **t**.: Nouns	*A*-score	Lexicosemantics (Poulin et al., 2016)
G. j.: Det.	*A*-score	Morphosyntax (Noonan et al., 2014)

[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 as the phonological and visuospatial working memory.

G. j.: Adj.	*A*-score	Morphosyntax (Noonan et al., 2014)
J. **t**.: Verbs	*A*-score	Lexicosemantics (Haebig et al., 2014)
G. j.: Regular	*A*-score	Morphosyntax (Noonan et al., 2014)
G. j.: Irr. agrm.	*A*-score	Morphosyntax (Noonan et al., 2014)
Corsi–Forward	Raw	Visuospatial working memory (Corsi, 1972; Baddeley et al., 2000)
Corsi–Backward	Raw	Visuospatial working memory (Corsi, 1972; Baddeley et al., 2000)
DMTS–1s	Raw	Visuospatial working memory (Daniel et al., 2016)
DMTS–5s	Raw	Visuospatial working memory (Daniel et al., 2016)

Note. Pcl = Percentile; Nonword rep. = Nonword repetition task; S-V = subject-verb; prod. = production; comp. = comprehension; G. j. = Grammaticality judgment; J. t. = Judgment task; Det. = Determiner; Adj. = Adjective; Reg. agrm. = Regular subject-verb agreement; Irr. agrm. = Irregular and subregular subject-verb agreement; DMTS = Delayed Match to Sample.

2. Supplementary Results

2.1. Variable selection

Twenty subtasks were administered, and they generated 26 scores. The variable selection analyses revealed that 17 had null Information Gain (IG). Significant differences were found in between groups on 14 tasks, when looking at the p-value of Brunner-Munzel test (Brunner & Munzel, 2000) without correction for multiple comparisons. See Table S2 for details.

Table S2. *Information gain, area under the curve and Brunner-Munzel tests for all 26 subtasks.*

Subtasks	IG	AUC	Brunner-Munzel tests		
			t_{bm}	p^a	CLES
Recalling Sentences – Percentile	0.52	0.99	–28.88	< .001	0.02
Word Classes Comprehension – Percentile	0.32	0.93	–10.8	< .001	0.07
fLEX Irregular Verb prod. – Raw score	0.27	0.92	–8.27	< .001	0.09
Recalling Sentences – Raw score	0.27	0.85	–5.55	< .001	0.15
EVIP – Percentile	0.25	0.89	–7.27	< .001	0.11
Nonword repetition – Raw score	0.24	0.85	–5.61	< .001	0.15
Forward Digit Span – Percentile	0.24	0.82	–4.34	< .001	0.18
Word Classes Production – Percentile	0.23	0.83	–4.17	< .001	0.18
Grammaticality Judgment: Irregular agrm. – A-score	0.21	0.83	–4.64	< .001	0.18
fLEX Irregular Verb comp. – Raw score	0	0.80	–4.21	< .001	0.2
Grammaticality Judgment: Regular agrm. – A-score	0	0.75	–2.91	.006	0.25
Forward Digit Span – Percentile	0	0.75	–2.99	.006	0.25
Word Classes Comprehension – Raw score	0	0.74	–2.65	.012	0.27
Word Classes Production. – Raw score	0	0.69	–2.06	.048	0.31
Backward Digit Span – Percentile	0	0.67	–1.82	.08	0.33
EVIP – Raw score	0	0.66	–1.6	.12	0.35
Grammaticality Judgment: Adjectives – A-score	0	0.65	–2.41	.026	0.34
Grammaticality Judgment: Verbs – A-score	0	0.62	–1.23	.227	0.38
Corsi–Backward – Raw score	0	0.61	–1.18	.246	0.39
Grammaticality Judgment: Nouns – A-score	0	0.61	–2.22	.041	0.38
Backward Digit Span – Raw score	0	0.59	–0.95	.349	0.41
fLEX action naming – Raw score	0	0.56	–0.78	.44	0.44
Grammaticality Judgment: Determiners – A-score	0	0.56	–1.46	.163	0.44

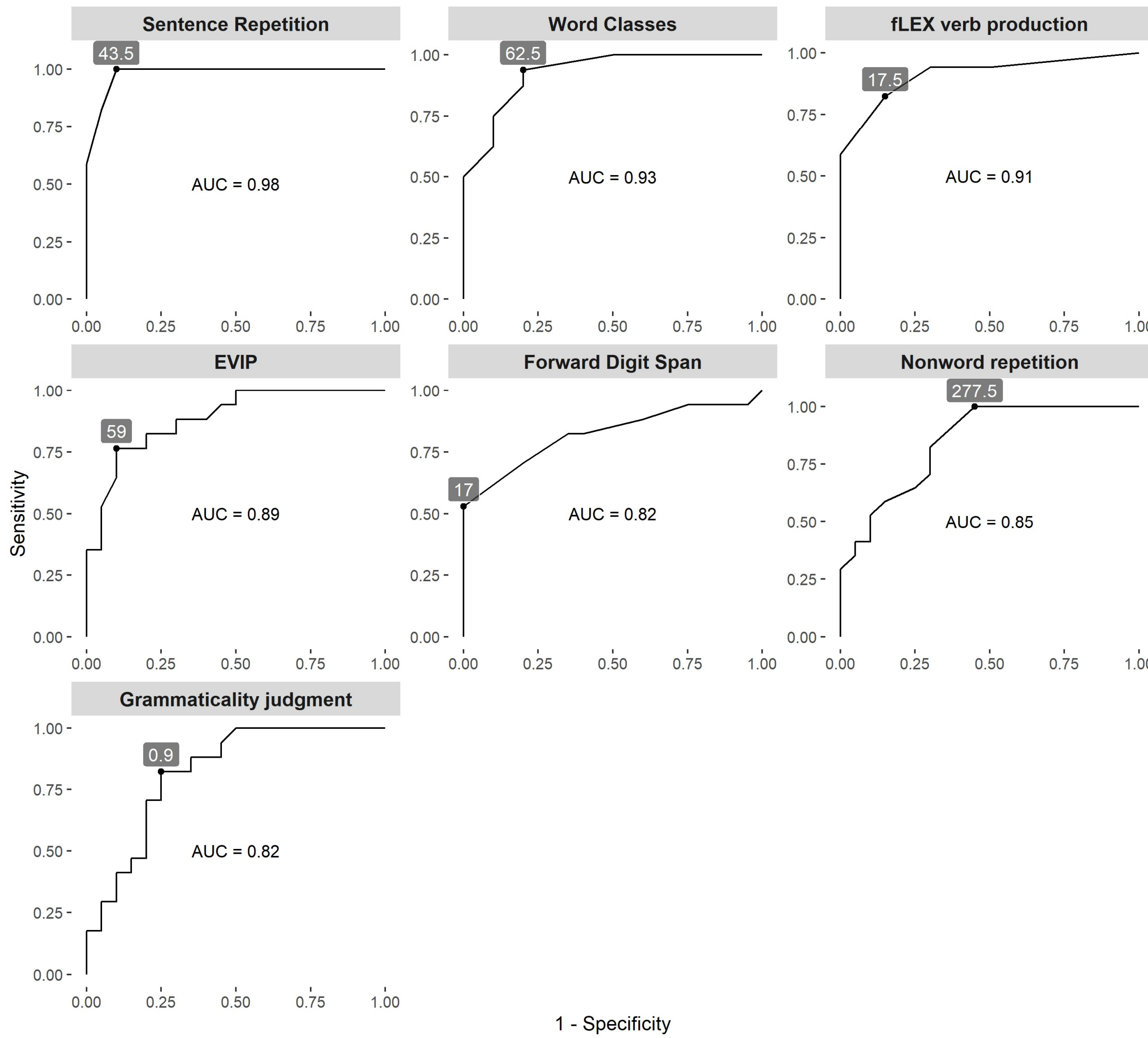
DMTS 1s – Raw score	0	0.55	–0.48	.631	0.45
Corsi Forward – Raw score	0	0.51	–0.12	.906	0.49
DMTS 5s – Raw score	0	0.47	0.24	.81	0.52

Note. IG. = Information gain; CLES. = Common-language effect sizes; AUC. = Area under the curve; agrm. = agreement; DMTS = Delayed Match to Sample. [a]The Brunner-Munzel *p*-values are presented with no adjustment for multiple comparisons.

2.2. Optimal cut-off scores

We present for the seven discriminating subtasks their ROC plot showing the area under the curve for all possible scores and with their respective optimal cut-off scores.

Figure S1. *ROC curves for all seven discriminating subtasks and their optimal cut-off scores.*



Note. Grey labels indicate optimal cut-off scores. Three score types are displayed: percentile scores (Recalling Sentences, Word Classes, Forward Digit Span and EVIP), raw scores (fLEX irregular verb production: maximum of 20 verbs, Nonword repetition: maximum of 280 phonemes) and one *A-score* (Grammaticality judgment of irregular verb agreement).

2.3. Multivariable analysis

We fitted a regularized logistic regression model, with group classification (TL or DLD) as a dependent variable, with our seven most discriminating subtask scores: Recalling Sentences, Word Classes, Forward Digit Span, EVIP, fLEX subject-verb agreement verb production and Grammaticality judgment of irregular and sub-regular verb agreement. When used to classify the participants between TL and DLD groups, the final model had a sensitivity of 0.94, a specificity of 0.95, and an AUC of 0.99. Of 36 participants with no missing values, the model accurately classified 34 and produced 1 false negative (one participant with DLD classified as being in the TL group) and 1 false positive (one participant with TL classified as being in the DLD group). Recalling Sentences subscores contributed the most to the model, followed by fLEX verb production. The other subscores did not contribute to the model.

Table S3. *Seven tasks' coefficients and variable importance for the regularized logistic regression model.*

Subtasks	Coefficient	Var. contribution (SD)
Recalling Sentences	–1.718	0.42 (0.093)
fLEX subject-verb verb production	–0.510	0.014 (0.013)
Word Classes	–	0 (0)
EVIP	–	0 (0)
Forward Digit Span	–	0 (0)
Nonword repetition	–	0 (0)
Grammaticality judgment irregular and subregular verbs	–	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 (AUC) when the variable is permuted; results shown are for 100 permutations.

References

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

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

Corsi, P. M. (1972). *Human memory and the medial temporal region of the brain (unpublished PhD thesis)*. McGill University, Montreal.

Courteau, É., Royle, P., Gascon, A., Marquis, A., Drury, J. E., & Steinhauer, K. (2013). Gender concord and semantic 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., Steinhauer, K. & Royle, P. (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

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

Duquette, A.-S., Courteau, E., & Royle, P. (2020, January 31). *Fidélité inter-juge d’une étude sur le trouble développemental du langage et de la compréhension du langage oral*. 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)

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

Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The children’s test of nonword repetition: A test of phonological working memory. *Memory*, *2*(2), 103–127. https://doi.org/10.1080/09658219408258940

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

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

Poulin, M.-J., Marquis, A., & Royle, P. (2016). Étude de faisabilité portant sur l’évaluation de la production et de la compréhension du langage oral en français. *ScriptUM: La Revue Du Colloque VocUM*. http://journal.vocum.ca/index.php/scriptum/article/view/28

Pourquié, M. (2017). Evaluation du lexique et de la flexion verbale dans le trouble primaire du langage. *ScriptUM : la revue du colloque VocUM*, *3*, 35–53. https://scriptum.vocum.ca/index.php/scriptum/article/view/47

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)*. The Psychological Corporation.

Van Kleeck, A. (1982). The emergence of linguistic awareness: A cognitive framework. *Merrill-Palmer Quarterly*, *28*(2), 237–265.