CAMBRIDGE
UNIVERSITY PRESS

ORIGINAL ARTICLE

Working memory training enhances complex syntax in children with Developmental Language Disorder

Hélène Delage[1]* , Emily Stanford[2] and Stephanie Durrleman[2]

[1]Equipe de Psycholinguistique et Logopédie, Faculté de Psychologie et des Sciences de l'Education, Université de Genève, 40 boulevard du Pont d'Arve, 1211 Genève, Switzerland and [2]Département de Linguistique, Faculté des Lettres, Université de Genève, 2 Rue de Candolle, 1211 Genève, Switzerland
*Corresponding author: Email: helene.delage@unige.ch

Abstract

Linguistic deficits attested in children with Developmental Language Disorder (DLD) have been explained in terms of limitations in working memory (WM). The goal of this research is to assess whether a tailored WM program can improve the syntactic abilities of children with DLD and those with typical development (TD). We created a novel iPad application consisting of five activities specifically designed to train the components of WM that have been shown to be the most predictive of performance on tests assessing complex syntax. Thirty-two children with DLD (M = 9;0) and 18 with TD (M = 8;5) followed the WM training (lasting 12 hours). Results show significant improvement in verbal WM (direct effects) in both TD and DLD groups, and in sentence repetition (transfer effects) in the DLD group, with the most pronounced improvements observed for complex syntactic structures. This progression is not observed for 38 age-matched children of the same age who followed an alternative, global scholastic training (20 DLD, 18 TD), which proves the specific efficacy of our WM training. The logical next step will be to incorporate the training into the therapy of children with DLD in order to reinforce the potential benefit of their interventions.

The terminology "Specific Language Impairment" (SLI), referring to children with oral language impairment for no obvious reason, was recently replaced by the term "Developmental Language Disorder" (DLD) (Bishop et al., 2017). This diagnosis, observed in 5–10% of the population (Law et al., 2019), includes children with severe and persistent language disorders that interfere with their communication in daily life and/or affect their schooling in the absence of a specific biomedical condition (such as Down syndrome). These two inclusion factors (severity and persistence) were already present in the original SLI terminology (Leonard, 2014), but it

was the exclusion criterion referring to the specificity of the disorder that was problematic. Indeed, the new terminology better reflects the clinical reality of comorbidity associated with DLD, as language deficits observed in DLD are often associated with difficulties in other realms, including executive functions such as sustained attention, working memory (WM), inhibition, and attention shifting (Ebert & Kohnert, 2011; Kapa & Plante, 2015; Kapa et al., 2017). Considering the various interconnections between language and other higher cognitive functions, Tomas and Vissers (2019) suggest that DLD be treated as a complex neuropsychological syndrome in which language delays would be explained by other neuropsychological deficits.

Among the deficits reported for this population are impairments in WM[1]. These weaknesses can be revealed via simple-span tasks that require the maintenance and recall of verbal information, such as nonword repetition, forward digit span, and serial word order span (Archibald & Harder Griebeling, 2016; Ellis Weismer et al., 1999; Gathercole & Baddeley, 1990; Hick et al., 2005; Montgomery & Evans, 2009), or via complex-span tasks that require processing while maintaining stored information, such as listening span, counting span, or backward digit span (Archibald & Gathercole, 2006; Delage & Frauenfelder, 2020; Hoffman & Gillam, 2004; Montgomery, 2000). These well-known WM deficits in DLD have been integrated into various theories accounting for the condition (Conti-Ramsden et al., 2015; Evans & Brown, 2016; Gillam & Hoffman, 2003; Kail, 1994; Nicolson & Fawcett 2007; Ullman & Pierpont, 2005).

Archibald and Gathercole (2006) revealed that impairments in WM observed in children with DLD (aged 7 to 11 years) in both simple and complex span tasks not only cooccur but even exceed the criterial language deficits that are characteristic of the disorder. Performance in visuospatial WM, on the other hand, gives more mixed results, with difficulties either not emerging at all (Baird et al., 2010; Ellis Weismer et al., 2017; Petruccelli et al., 2012) or being restricted to only subgroups of children with DLD (Archibald & Gathercole, 2006; Henry & Botting, 2017; Vugs et al., 2014). The observations that limitations in verbal (but not necessarily visual) WM are common to the majority of children with DLD could help to explain their language deficits. In line with this view, Baird and colleagues (2010) assessed the WM performance of three groups of children aged 6 to 16 years and revealed significant verbal WM deficits in the first group with DLD (N = 51), more subtle WM deficits in the second group with a history of resolved language impairment (N = 13), and an absence of WM deficits in the third group with no history of language impairment (N = 26).

Working memory and language in TD and DLD

The relation between WM and language has been extensively studied, both in typical and atypical language development, as well as in second language acquisition (e.g., Mackey and Sachs, 2012). In typical development (TD), research has established clear links between WM and lexical acquisition (Leclercq & Majerus, 2010; Majerus et al., 2006) as well as between WM and syntax in school-aged children (Adams & Gathercole, 2000; De Abreu et al., 2011; Delage & Frauenfelder,

2019; Ellis Weismer et al., 2017; Finney et al., 2014; Montgomery et al., 2008; Poll et al., 2013; Willis & Gathercole, 2001). These associations have been highlighted with simple span tasks (e.g., nonword repetition tasks and forward digit span in children aged 4–5 years, Willis & Gathercole 2001) as well as with complex span tasks (e.g., listening span task in children aged 6–13 years, Poll et al., 2013). In a study combining the two types of span, Delage and Frauenfelder (2019) found a predictive relation between the two composite WM scores (simple and complex spans) and different measures of syntax: repetition, comprehension, and spontaneous production of complex sentences, in 48 TD children aged 6–12 years.

The significant difficulties with WM experienced by children with DLD persist into adolescence (Ellis Weismer et al., 2005), as do their language deficits (see Conti-Ramsden & Durkin, 2008), particularly for complex syntax (Nippold et al., 2009; Tuller et al., 2011, 2012). Such syntactic difficulties are accounted for by the Derivational Complexity Hypothesis (DCH, Jakubowicz, 2011; Jakubowicz & Strik, 2008; Jakubowicz & Tuller, 2008)[2]. This metric predicts that children with DLD should encounter difficulties with noncanonical sentences, which have been borne out by studies on object clitic pronouns (see, for example, Tuller et al., 2011), object relatives (1) (Adani et al., 2016; Delage & Frauenfelder, 2020), object questions (2) (Friedmann & Novogrodsky, 2011; Jakubowicz, 2011), and passive sentences (3) (Marinis & Saddy, 2013; Montgomery & Evans, 2009). All these different constructions share the underlying property of implying a noncanonical word order instead of the French default order: subject–verb–object. More specifically, in all examples, the (logical) object moves to a higher position, a syntactic movement symbolized by the arrows and resulting in different word orders.

(1) Montre-moi la pomme que Mary mange __   OSV

Show me the apple that Mary eats

'Show me the apple that Mary is eating'

(2) Quelle pomme Mary mange __ ?   OSV

Which apple Mary eats

'Which apple is Mary eating?'

(3) La pomme est mangée __ (par Mary)   OVS

The apple is eaten (by Mary)

'The apple is eaten (by Mary)'

The DCH also links mastery of grammatical structures to cognitive factors: "*Jakubowicz (2004, 2005) proposed that (ab)normal language build up is affected by developmental constraints such as the capacity of working memory, that are sensitive to the computational complexity of the derivation*" (Jakubowicz & Strik, 2008:

106). More computational complexity would be involved in the structures above, where the movement operation would lead to an overload of WM capabilities. The immature WM systems of young TD children would thus yield a grammar characterized by the presence of short and simple, canonical sentences. In children and adolescents with DLD, limitations in WM persist, potentially explaining reports of their avoidance of noncanonical structures (see above) and links between performance on these complex structures with WM.

To take the concrete example of an object relative clause, its mastery requires 1) remembering the object until reaching the position of gap, i.e., the trace/copy left by the moved element, 2) while processing the new verbal stimuli (= words) to be integrated into the sentence, and then 3) proceeding to the correct interpretation of the sentence by linking the verb to its different arguments (Gibson, 1998). WM resources are thus plausibly solicited by such structures, which would therefore pose challenges to children with a condition displaying memory deficits. This reasoning has found empirical support from a series of studies. For instance, in young TD children, Arosio and colleagues (2011) showed that digit span performance relates to the comprehension of object relative clauses. Similarly, Bentea and colleagues (2016) report that WM measures (assessed by forward and backward digit spans) correlated with the comprehension of object Wh-questions and object relative clauses in French-speaking children aged 5 years. In older children (aged 7 to 9 years), this link was limited to the most complex constructions (involving intervention effects in the presence of two NPs[3]). Still on object questions and object relative clauses, Stanford and Delage (2019) found that comprehension of English-speaking children with specific learning disorder (mainly with dyslexia) aged 7 to 11 years correlated with their performance on backward digit span. Finally, Frizelle and Fletcher (2015) have also identified a strong relationship between the repetition of relative clauses, varying in complexity, and WM measures in 35 English-speaking children with DLD aged 6–7 years. More precisely, complex-span tasks (including listening recall, counting recall, and backward digit span) were particularly associated with the more complex relative clauses, like object relatives. In a similar study, Riches and colleagues (2010) examined several WM measures (nonword repetition, digit span, and listening span) as well as the capacity to repeat relative clauses in adolescents with DLD aged 14 to 16 years. Their results reveal that performance on repetition of such complex sentences was significantly correlated to their scores obtained on WM tasks.

Other constructions involving syntactic movement, such as passives, were also found to be linked to WM in 25 English-speaking TD children aged 7 years with a listening recall task (Marinis & Saddy, 2013) and in adults with a composite measure of WM-capacity index (Sung et al., 2017). Mastery of such complex grammatical constructions arguably requires storing and manipulating verbal sequences since "*these structures require storing of the NPs of the sentence in memory before syntactically and semantically integrating with the verb phrase thanks to the cue provided by the passive morphology*" (Durrleman et al., 2017: 8). Montgomery and colleagues (2008) compared the role of simple spans (assessed via nonword repetition) and complex spans (assessed via listening span) in the comprehension of a variety of both simple and complex sentences in children aged from 6 to 12 years. Complex structures assessed were passives or sentences involving binding

dependencies (with either reflexive or accusative pronouns), while simple sentences contained no movement or dependency. A correlation emerged between the comprehension of complex sentences and performance on complex spans, but no correlation was observed between the comprehension of simple sentences and simple spans. Furthermore, regression analyses showed that complex-span results explained 30% of the variance in the comprehension of complex sentences. These findings were replicated for children with DLD (Montgomery & Evans, 2009).

Syntactic movement is not the only source of syntactic complexity for children with DLD. These children also have difficulty with complex sentences that include embedding (Delage et al., 2008; Hamann & Tuller, 2014; Scheidnes & Tuller, 2014). For example, Tuller et al. (2012) show, through a spontaneous language analysis conducted in 18 adolescents with DLD aged 11 to 16 years, that they produced shorter sentences with fewer embedded sentences than 8-year-old TD children. Few studies have focused on this aspect with respect to WM. However, in young TD children aged 3–5 years, Adams and Gathercole (2000) and Willis and Gathercole (2001) have shown that children with better WM skills (assessed by word, nonword, and number repetition tasks) produced and repeated longer and more complex sentences than children with poorer memory skills. Mastering the production of complex sentences requires dealing with structures in which the different verbs must be related to several arguments and which include tense concordance phenomena, while keeping in mind the previous words of the sentence. In the case of multiple embedding as in (4), the implication of WM capabilities appears even more obvious, as Kimball (1973) postulated that the number of incomplete clauses that must be stored in memory in these contexts crucially influences processing load.

(4) Mary croit [que son fils préfère la maîtresse [qui est plus jeune]].
    'Mary believes that her son prefers the teacher who is younger'

Tuller et al. (2012: 165) explicitly linked both the level of embedding and the number of operations involved in a derivation to WM: "These two properties both appear to be related to demands on memory: building syntactic structure while keeping in mind already built syntactic structure (particularly when a new clause is involved), linking antecedents and gaps, comparing features for agreement". This approach is in line with the Jakubowics's DCH approach previously mentioned which aims to account for the syntactic impairment in DLD by linking domain-specific syntactic principles with domain-general behavioral variables. Even if these approaches are firmly grounded in the nativist tradition, they acknowledge that cognitive capacities, and especially WM, play a role in language acquisition, and more precisely that they are recruited to help develop domain-specific language predispositions. Chomsky himself acknowledged this (2005: 12): "It could be that unbounded Merge, and whatever else is involved in [Universal Grammar], is present at once, but only manifested in limited ways for extraneous reasons (memory and attention limitations and the like)." In more purely cognitive approaches to language acquisition, such as usage-based models (e.g., Tomasello, 2000, 2005, 2009), language acquisition reflects the maturation of cognitive capacities and processes. From this point of view, children are seen as using their general and

social-cognitive skills to build up an inventory of linguistic constructions through imitation of the language(s) they hear around them (Tomasello, 2000). Since language acquisition is assumed to interact closely with the gradual development of other cognitive processes within this framework, limitations in cognitive skills, such as WM, should play a nonnegligible role in language outcomes.

Tuller and colleagues' hypothesis was addressed by Zebib et al. (2019) who explored the performance of 23 bilingual children (6–8 years old) with DLD and 53 bilingual TD children of the same age in sentence repetition (with complex sentences varying in embedding and syntactic movement) and WM, using classical tasks of forward and backward digit spans. Results showed that sentence repetition was mainly linked to complex spans in DLD children, whereas it was predicted by simple spans in TD children, suggesting that the former rely more on their general processing abilities to repeat complex sentences that they do not yet master. In previous work that provided the basis for the current study, Delage and Frauenfelder (2020) also addressed these two complexity factors of embeddedness and syntactic movement by using more diverse syntax and memory tasks. Twenty-eight children with DLD (5–14 years old) and 48 TD children of the same age, all monolingual and francophone, were assessed for verbal WM skills through simple span tests (forward digit-span, serial memory, and nonword repetition) and complex span ones (backward digit span, counting span and running span), and for syntax via tests of comprehension and repetition of complex sentences, as well as via spontaneous language samples. Results confirmed the persistent deficits of the children with DLD in all tests of WM and complex syntax and, most importantly, a strong relationship between WM and syntactic complexity, even though differences in age and nonverbal intelligence were controlled for. More specifically, results for the children with DLD revealed a strong predictive link between the serial component of verbal WM and syntactic abilities, a relation that is not mediated by more general intellectual skills. In TD children, the involvement of complex spans (including counting span and running span) appeared stronger than that of simple spans to explain the variance of the results in complex syntax. The clinical implications of these results have given rise to the current investigation, which aims to evaluate the specific effects of training WM to enhance grammatical abilities in DLD. Indeed, given that the WM limitations observed in children with DLD predict their difficulties in complex syntax, integrating WM training in speech and language therapy for this population may be promising.

Working memory training

In 2011, Morrison and Chein showed that WM capacity can be expanded through targeted training, involving the repetition of tasks on specific mechanisms. Several studies have since examined the effects of WM training, most often with healthy adults, yielding variable results: they generally show an overall improvement in WM performance without necessarily long-term maintenance and very little transfer to other cognitive abilities (verbal abilities, identification of written words, etc., see Melby-Lervåg & Hulme, 2013). The internal validity of these studies has been called into question by numerous scholars, who have thus encouraged the scientific

community to conduct future research using more appropriate experimental designs and measures adapted to large cohorts of participants (Majerus, 2016; Melby-Lervåg & Hulme, 2013, 2016). Interestingly, when WM training focuses on populations with learning disabilities, results seem more promising, with short and long-term effects on WM and word decoding, as shown by the meta-analysis of Peijnenborgh and colleagues (2015). However, while WM training has been offered to various populations such as children with attention deficit hyperactivity disorder (ADHD) or learning disorders (Majerus, 2016), very few studies have focused on children with oral language difficulties. One that did is Holmes and colleagues (2015) which reported improved visuospatial skills in 12 children with poor language skills aged 8 to 11 years, following WM training by means of the *Cogmed* program. Vugs and colleagues (2017) also proposed executive function training (including visuospatial WM, inhibition, and cognitive flexibility tasks) to 10 children with DLD aged 8 to 12 years. The authors observed a gain that seemed to be maintained on executive functions but did not test for possible language progression, while recognizing that: "If it is proven possible to improve [executive functions] in children with SLI by computerized training, this might also have an (indirect) effect on the linguistic skills of these children" (p. 16). Ebert and Kohnert (2009) examined the language effects of processing speed and auditory memory training in two children with DLD aged 7 and 8 years. Results indicated that the participants made gains in processing speed and language abilities, including sentence formulation and grammatical morpheme production. Note, however, that such case studies do not allow the results to be generalized to a population as heterogeneous as DLD.

Aim and prediction of the present study

The current study focused on children with DLD and proposed a targeted training of verbal WM. The intensive training programs for WM that already exist on the market (e.g., the *Cogmed* or *Jungle Memory* programs) seem too broad to enhance the specific skills identified as underlying linguistic complexity, as they include a number of unrelated activities such as visuo-spatial memory. We thus developed *Magic Memory* (Delage et al., 2017), a program aimed at training those precise aspects of WM that have been shown to predict performance in complex syntax, i.e., complex spans and serial memory (Delage & Frauenfelder, 2019, 2020). This program has shown its effectiveness on the ability of children with DLD to produce accusative clitic pronouns (Stanford et al., 2019), structures that, it should be recalled, include syntactic movement. Specifically, 26 children with DLD aged 5 to 12 years who had completed the WM training program were compared to 17 children with DLD receiving an alternative (scholastic) training. The results showed that after 12 hours of effective training in WM, the first group made significant progress in the production of object clitics, while the performance of the control group remained stable. Furthermore, the program had direct effects with an improvement in WM capacities in children with DLD. These direct effects were also observed in 16 children of the same age with typical language development.

The objective of the present research is to extend the cohort of DLD and TD children benefiting from intensive WM training and to explore whether transfer

effects can be found for syntactic structures varying in the number of (i) syntactic derivations and (ii) embedded clauses. Whereas our previous study aimed at eliciting a single syntactic structure (i.e., accusative clitics), we propose here to explore the repetition of varied syntactic structures, consisting of simple sentences and relative clauses, for which the factors "syntactic movement" and "embedding" have been conscientiously manipulated[4]. As it was the case for elicitation tasks of accusative clitics in romance language, sentence repetition tasks have been shown to be good clinical markers of DLD (e.g., Archibald & Joanisse, 2009; Armon-Lotem & Meir, 2016; Zebib et al., 2019), even in young children (Everitt et al., 2013). This type of task assesses multiple components such as expressive phonology, lexical knowledge, morphosyntactic skills, as well as verbal WM (Alloway & Gathercole, 2005; Polišenská et al., 2015; Tuller et al., 2018). To focus on syntax and then to minimize the impact of the other skills, we proceeded to different adjustments:

- We contrasted simple and complex sentences that contained the same number of syllables, to counterbalance the effects of verbal WM;
- We considered measures for which children respect or not the expected structure and level of embedding, whatever the lexicon they used, to neutralize effects of lexical selection/retrieval;
- We did not penalize any phonological deformations, to overcome the effects of imprecise phonological representations or articulatory deformations.

More precisely, our predictions are as follows:

1. **As in our previous study, we expect that children (both TD and DLD) following the WM training will improve their capacities in WM, i.e., direct effects, and these improvements will be observed in tasks beyond those trained by the program**. As for children following the control training, we predict no significant improvement in WM, which would prove the specificity of our WM training. These effects have already been found in our earlier study, but we think that replicating them with a larger cohort of participants would strengthen our conclusions and confirm the effectiveness of our program. This seems to be of particular importance as such effectiveness of WM training is still being debated in the literature (Melby-Lervag & Hulme, 2013; Melby-Lervag et al., 2016). Moreover, we aim to explore another aspect not previously addressed, namely, whether the observed effects can be sustained over time through delayed posttest.

2. Next, and more importantly, we also expect to demonstrate the presence of **transfer effects**, with children trained in WM also improving their ability to accurately repeat complex sentences?. Here again, we predict that improvements in syntax will not be observed in children in the control (scholastic) group who would not have undergone any specific training to improve this area. As for differences between TD and DLD children in the WM training, we expect a modest progression in the former compared to the latter, since the syntactic level of TD children would be logically higher than that of DLD children, leaving a smaller margin for progression.

Table 1. Characteristics of experimental (WM) and control (Scholastic) groups

Cognitive group	Type of training	Groups	N	Gender	Age range	Mean age	N simultaneous bilinguals
DLD	WM	DLD$_{WM}$	32	8f, 24m	6;5-12;5	9;0 (1;6)	10
	Scholastic	DLD$_{SQ}$	20	6f, 14m	6;0-11;9	9;3 (1;9)	2
TD	WM	TD$_{WM}$	18	10f, 8m	6;0-12;4	8;5 (2;0)	1
	Scholastic	TD$_{SQ}$	18	7f, 11m	6;0-12;7	8;4 (1;11)	1

3. Finally, we will be able to explore to which extent transfer effects are **modulated by syntactic complexity factors**, given our manipulation of the degree of complexity of the structures to be produced. We predict that the performance of children in the WM training group will improve for the most complex structures, which are hypothesized to particularly tax WM resources, and that this effect will be less obvious for simpler syntactic structures.

Method

Participants and inclusion criteria

A total of 88 native French-speaking children (52 with DLD and 36 with TD) aged 6 to 12 years took part in this study. This age range was chosen since it corresponds to the ages of children for whom a predictive relationship between WM and syntax had previously been identified (Delage & Frauenfelder, 2019, 2020). Thirty-two children with DLD ('DLD$_{WM}$', *M* age = 9;0) followed the WM training entitled *Magic Memory* (MM) and were compared to 20 children with DLD matched on age ('DLD$_{SQ}$', *M* = 9;3) who followed an alternative, scholastic, training program called *Squla*. TD children were also divided between WM training ('TD$_{WM}$', N = 18, *M* age = 8;5) and alternative training ('TD$_{SQ}$', N = 18, *M* age = 8;4). Assignment to the different training groups was random. Even though age matching was only done for each cognitive group (DLD, TD), the four groups did not significantly differ for age, $F(3, 84) = 1.3$, $p = .28$. Table 1 details the general characteristics of the different groups (gender, age range, mean age, and bilingualism). As for inclusion criteria, we only recruited monolingual (N = 74) or simultaneous bilingual (N = 14) French-speaking children, hence only native speakers of French participated.

Cognitive groups were based on children having received a diagnosis of language impairment by speech and language therapists, hence being assigned to the DLD group, or an absence thereof for TD children. More specifically, children with DLD had been tested by clinicians with standardized tasks (e.g., the EXALANG 5-8, Thibaut et al., 2010), and we required that these demonstrate clear syntactic impairment, since DLD does not always include such deficits (Friedman & Novrogrodsky, 2008). We also made sure with the speech-language therapists that these children did not present any known differentiating condition (as defined by the CATALISE group, Bishop et al., 2017), such as brain injury, aphasia, cerebral palsy, sensorineural hearing loss, autism spectrum disorder, or intellectual disability. TD children had normal levels of language (i.e., no language therapy) and were

Table 2. Standardized assessments: nonverbal reasoning, expressive grammar, and working memory

			Working memory	
Group	Nonverbal reasoning	Expressive grammar	Simple spans	Complex spans
DLD$_{WM}$	−0.4 (0.8)	−3.2 (2)	−1.3 (1.1)	−2.1 (0.8)
DLD$_{SQ}$	−0.5 (0.7)	−3.3 (1.8)	−1.7 (0.8)	−2.4 (0.6)
TD$_{WM}$	0.6 (0.7)	0.9 (1.2)	0.8 (0.8)	−0.8 (0.7)
TD$_{SQ}$	0.8 (0.5)	1.2 (1.1)	0.9 (0.7)	−0.6 (1.0)

functioning without special difficulties in their age-appropriate classroom, according to their parents. Different standardized tests were also used to refine our inclusion criteria. First, we ensured that both groups performed above the 10th percentile in a nonverbal reasoning task (Raven's Colored Progressive Matrices, Raven & Court, 1998), thereby excluding any risk of intellectual disability. Then, we conducted a standardized assessment of expressive grammar for all children with a subtest of the BILO-3C battery (Khomsi et al., 2007) as well as of WM via a battery which combined three simple-span tasks and three complex-span tasks (Boutard & Gatignol, 2015). To be definitely included in the study, children with DLD had to obtain a score of at least 1.25 SD below age-specific norms for the score in grammar as well as such delay for a minimum of three of the six WM tasks[5]. As for TD children, we only retained participants who scored with a score above −1 SD in expressive grammar and had no more than two of the six WM scores that were inferior to −1 SD[6]. In this manner, we ensure that children with DLD had clear deficits in WM and in syntax, whereas TD children showed preserved capacities. Table 2 presents the mean standard deviations obtained by the different groups for these standardized tasks. Two composite scores were calculated for WM tasks, grouping simple-span, and complex-span tasks. Using independent t-tests, the two groups of DLD children (DLD$_{WM}$ versus DLD$_{SQ}$) did not significantly differ for nonverbal reasoning ($p = .65$), expressive grammar ($p = .75$) or WM, whether for simple ($p = .11$) or complex spans ($p = .09$)[7]. The same pattern was observed for TD children (respectively $p = .35$, $p = .41$, $p = .77$, $p = .37$). As expected, children with DLD, as a whole, differed significantly from TD children for expressive grammar, $t(86) = −12.2$, $p < .001$, $d = 2.76$, and for WM measures, for simple-span, $t(86) = −11.3$, $p < .001$, $d = 2.52$, and complex-span tasks, $t(86) = −8.3$, $p < .001$, $d = 1.7$. These two groups also differed for nonverbal reasoning, $t(86) = −7.4$, $p < .001$, $d = 1.63$, with DLD children scoring lower than TD ones. This difference can be explained by the fact that children with DLD, even though they display performance within the norm, typically perform lower than age-matched controls on nonverbal tasks (Leonard, 2014).

Of the 14 bilingual children, 12 have a diagnosis of DLD. We compared the performance of these 12 children to that of the other (monolingual) children with DLD (N = 40). These two groups did not significantly differ for age ($p = .50$), nor for expressive grammar ($p = .08$), nonverbal reasoning ($p = .05$) or complex-span tasks ($p = .46$). The only difference which emerged was in simple-span task where

bilingual children outperformed monolinguals, $t(50) = 3.0$, $p = .004$, $d = 1.03$. Note, however, that the effect size is reduced, as is the sample of bilinguals, and that the difference seems to be due to the particularly high performance of some bilingual children (see also Marini et al., 2019 for similar effects in bilinguals). Ages and standardized scores of monolingual and bilingual children with DLD are provided in Appendix A.

Material and methodology

General procedure

For all of the trained participants, we established a baseline via pretests, conducted one week before the first training session. These tests evaluated various WM and syntactic skills. One week after the last training session, posttests evaluated these same skills using different (but matched) items, in order to avoid a learning effect. Hence, we obtained two different test versions, A and B, in which words were controlled and matched for length end frequency. The order of these versions was randomized between participants in order to have half of the participants passing version A in pretests and version B in posttests, and the opposite for the other half. These measures are used to identify potential progression of trained participants and thus to assess the immediate effects of training. Three months after the posttests, we conducted another session of posttests with DLD participants (N = 12) who had followed WM training, who were available at this moment and whose parents agreed to participate in this third testing phase. For these delayed posttests, participants were readministered the version (A or B) they had completed at the pretest, 5 to 6 months earlier, which was sufficient to avoid potential retest effects.

The training sessions consisted of two different training programs. As previously mentioned, the target "*Magic Memory*" (MM) training aimed to improve WM capacities with different exercises implying simple and complex verbal spans. The alternative training, entitled "Squla," focused on scholastic skills with exercises adapted to the participants' school level. The aim of this control training was to ensure that any potential gains in WM and/or syntax following the training period were indeed related specifically to the WM training program rather than to maturation or global cognitive stimulation. More specifically, we predicted that the progression of the group benefiting from WM training would be better, both in WM and syntax, than that of children who had followed the alternative training. Both training sessions were offered with the same duration and format: each participant completed three training sessions per week, with each session lasting 30 minutes, for 8 weeks, which constitutes a total of 12 hours per participant. Both programs were provided in computerized format, either on iPads or computers, depending on the material available at children's homes. For each activity of both types of training, the task difficulty was adapted to match the child's performance level. Frequent positive feedback was also included in the programs, in order to boost children's motivation. Figure 1 outlines the overall experimental design of the study.

The training sessions were carried out at the home of the participant under the supervision of parents and graduate students in French-speaking Switzerland, as well as in France. Students contacted parents on a weekly basis to ensure that the assigned training program was being appropriately followed and visited participants two to
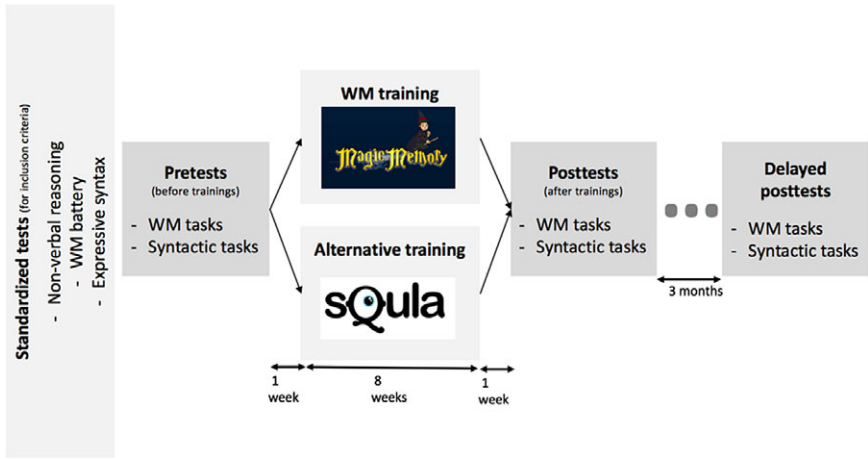
Figure 1. Experimental design.

three times a month to track the progress of the training regime. Approval for the research was obtained from the Ethics Committee of the University of Geneva and was also declared at 'La Commission Nationale de l'Informatique et des Libertés (CNIL)' in France. All parents received detailed information on the study and signed the consent form to approve the participation of their child.

Pre- and post-training tests: Working memory

Before and after the training phase, children passed a series of tests assessing WM with three tasks for simple spans and two tasks for complex spans (Table 3). As said before, we used two paired versions (A and B) for each task. Standard scores were derived for the individual tests and composite scores were calculated by averaging standard scores for each set of tests (simple-span and complex-span tasks).

Pre- and post-training tests: Syntax

Sentence repetition task
To test the ability of our participants to repeat syntactically complex sentences, we adapted a sentence repetition task created by Delage and Frauenfelder (2019)[8] that required participants to immediately repeat sentences read to them by the experimenter. The task was composed of 23 sentences: 8 syntactically simple control sentences and 15 syntactically complex experimental sentences. All sentences, both control and experimental, contained 14 syllables. As for experimental sentences, syntactic complexity was measured by the number and type of syntactic movement operations required to derive the structure (short- and long-distance phrasal movement or head movement) and the number of embedded clauses within the structure (1, 2, or 3), as shown in Table 4. There were thus simple sentences, without any embedding, which were intercalated (one simple sentence after two complex

Table 3. Pre- and post-training tests: Working memory tasks

	Tasks	Description	Score used
Simple span tasks	**Forward digit recall** (WISC IV, Wechsler, 2005)	The experimenter says aloud a series of digits increasing in length from 2 to 9; participants have to immediately repeat them aloud in the same order. Testing is discontinued when participants fail two trials in a row.	N correctly repeated sequences
	Nonword repetition (BELEC, Mousty et al., 1994)	The experimenter says aloud a nonword, which the participant must repeat immediately. Words increase in length from 1 to 5 syllables and in phonological complexity (with Consonant–Vowel and Consonant–Vowel–Consonant structures), such as *moga*, *juséga*, or *kragrinblan*. There is no stop criterion. A trial is marked as incorrect when (participants omit, add, or misorder one phoneme.	N correctly repeated syllables
	Serial-order word span (Majerus, 2008)	This task, presented as a game, tests specifically the ability of participants to retain serial-order information. They were required to store and recall only the serial order of items but not the names of items themselves. The children listened to sequences that contain familiar animal names along with the order in which these animals finished in a race. They had to place the cards corresponding to the animals on a winner's podium. The length of the sequences to be retained increased from two to seven animals depending on the children's performance.	N items retrieved in the correct order
Complex span tasks	**Backward digit recall** (WISC IV, Weschler, 2005)	The experimenter says aloud a series of digits increasing in length from 2 to 9; participants have to immediately repeat them aloud in reversed order. Testing is discontinued when (participants fail two trials in a row.	N correctly repeated sequences
	Counting span (Case et al., 1982)	After checking the capacity of children to count collections of up to 11 items, we asked them to count the number of blue dots on each page, while remembering the number of dots they had counted on the previous page(s). When a smiley appeared, they recalled the different numbers of dots, in the order of presentation. The number of pages increases until a stop criterion was reached.	N digits retrieved in the correct order

sentences) with complex sentences. The first relative clauses only contained one degree of embedding, followed by sentences with two and three degrees of embedding. For sentences with one degree of embedding[9], there were three subject

Table 4. Sentence repetition task

Simple sentences without movement or embedding	
Simple sentence (SVO), n = 8	Le petit garçon va à la piscine avec son frère *The little boy is going to the swimming pool with his brother*
Complex sentences varying in target structure	
Subject relative (SVO), n = 5	La maîtresse voit le garçon qui lit un livre sur Noël *The teacher sees the boy who is reading a book about Christmas*
Object relative (OSV), n = 5	Voilà une petite fille que je connais depuis longtemps *There is a little girl that I have known for a long time*
Object relative with inversion (OVS), n = 5	C'est un chat que caressent tous les enfants après l'école *It's a cat that all of the children pet after school*
Complex sentences varying in the number of embedded clauses	
1 degree of embedding, n = 9	La maîtresse voit le garçon [qui lit un livre sur Noël] *The teacher sees the boy [who is reading a book about Christmas]*
2 degrees of embedding, n = 3	Je crois [que la fille préfère le chien [qu'elle a colorié]] *I think [that the girl prefers the dog [that she colored]]*
3 degrees of embedding, n = 3	Il pense [qu'elle dit [que le garçon déteste la fille [qui pleure]]] *He thinks [that she says [that the boy hates the girl [who is crying]]]*

relatives and six object relatives (3 without subject−verb inversion, 3 with such inversion). For the second and third levels of embedding, there was one subject relative and two object relatives (1 without subject−verb inversion, 1 with such inversion) for each level. Appendix B presents the set of sentences for the two versions (A and B), which were matched for length, syntactic structure and frequency.

As the task progressed, the structures the children were asked to repeat became increasingly more complex. Three measures were considered: 1) the number of syllables that were correctly repeated, without considering the phonetic deformations, 2) the respect of the target structure (e.g., an object relative), and 3) the respect of the expected degree of embedding (with one, two, or three embeddings). For these two last measures, only the syntactic properties are considered; hence, the child obtained the point if s/he produced the correct structure and/or the expected level of embedding even if s/he changed the lexicon. Appendix C provides examples illustrating this scoring. After initial transcription and coding, all transcriptions were checked, and corrected if necessary, by two different experts.

Working memory training: Magic Memory

For the target WM training, we created a new training program called *Magic Memory*. This is an adaptive type of training where the level of difficulty gradually increases according to the progress made by the participants. Figure 2 (screenshots) illustrates
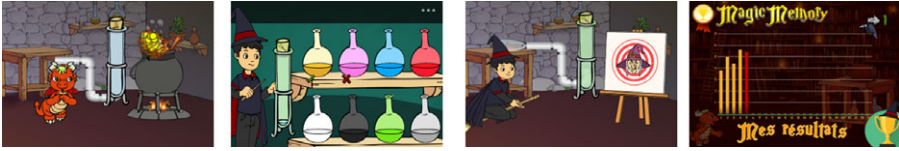
Figure 2. *Magic Memory*: Global design and examples of feedbacks.

the design of the program and the different types of feedback. In the *Magic Memory* training program, five activities were designed to train serial-order memory, WM updating, as well as dual-task processing. For the latter, the child must retain the order of familiar auditory stimuli while simultaneously performing a secondary task (a visual comparison of quantity task). The different activities, which were integrated and proposed in random order at each session, are detailed in Appendix D.

Alternative scholastic training

The alternative training program was offered through the *Squla* online educational game platform (Squla Inc. 2017). This training program for scholastic skills was carefully selected from a variety of existing systems. The program targets scholastic skills, such as spelling, history-geography, English, or mathematical reasoning, using several educational games with multiple choice questions and playful reinforcers. We made sure that none of the proposed activities focused heavily on WM or complex syntax. In addition, we asked parents to vary the academic notions to be worked on. The activities were adapted to each child's grade level. Trophies and positive feedback punctuated the different activities.

Results

Pretraining tests: Preliminary analyses

Before investigating whether or not our WM training program had been effective, we wanted to first ensure that there was a difference between our two cognitive groups (TD vs. DLD) but not between our two training groups ($_{WM}$ versus $_{SQ}$) on our three main pretest measures: (i) composite simple span score, (ii) composite complex span score, and (iii) percentage of correctly repeated syllables in the sentence repetition task, see Table 5. Independent *t*-tests confirmed that the performance of our TD participants was indeed significantly better than that of our participants with DLD on our pretest measures (TD$_{WM/SQ}$ children outperformed DLD$_{WM/SQ}$ children) and that our two training groups contained participants with similar pretest abilities (DLD/TD$_{WM}$ children performed similarly to DLD/TD$_{SQ}$ children).

As for the sentence repetition task in particular, we proceeded to an examination of unexpected structures and error types which showed that 1) the dominant non-expected structure consisted of producing simple sentences, without any embedding, instead of the target relative clause; 2) gender errors and omission/substitution of complementizers were the most frequent errors in both populations (TD/DLD). A more exhaustive description is provided in Appendix E.

Table 5. Summary of composite WM scores and pretest syntax scores

	Simple span		Complex span		Sentence repetition	
	M (%)(SD)	comparison	M (%)(SD)	comparison	M (%)(SD)	comparison
DLD	37.77 (8.98)	***	22.71 (8.55)	***	65.36 (15.87)	***
TD	56.14 (11.71)		40.48 (14.91)		91.98 (7.13)	
DLD$_{WM}$	37.92 (9.21)	ns	23.34 (7.40)	ns	66.67 (15.50)	ns
DLD$_{SQ}$	37.53 (8.84)		21.72 (10.26)		64.05 (16.24)	
TD$_{WM}$	57.33 (12.89)	ns	39.79 (13.81)	ns	90.21 (10.43)	ns
TD$_{SQ}$	54.96 (10.65)		41.16 (16.31)		93.76 (3.83)	

***$p < .001$

Direct effects: Is there an improvement in WM tasks?

After verifying that variables met standard assumptions of normality and heterogeneity, we wanted to examine whether the individual training programs had differential effects on the WM tasks that had not been directly trained[10]. In other words, was there an improvement from pre to posttest and if so, was this specific to the type of training program the participants had followed ($_{WM}$ versus $_{SQ}$)? To begin with, repeated measures ANOVAs were run with time (pretest, posttest) as the within subjects variable and training type (WM, SQ) and cognitive group (DLD, TD) as the between subjects factors. Our first analyses used composite scores for simple and complex span and revealed a main effect of time (posttest > pretest) and a main effect of cognitive group (TD > DLD) for both measures, but no main effect of type of training for either simple or complex span, see Table 6. In other words, when composite scores were considered, the WM and SQ groups (TD + DLD) performed similarly, in both the pre and posttest phases.

When the WM tests were analyzed individually, significant main effects of time (posttest > pretest) and cognitive group (TD > DLD) were observed for all five tests. In addition, for serial order word span there was a significant main effect of type of training, with children in the WM training group outperforming children in the SQ training group. For the other four tests, this effect was not significant, see Table 7.

Next, the repeated measures ANOVAs were used to investigate the effects that were of primary interest to us, mainly if the target training had a specific effect across the two cognitive groups (DLD, TD) on WM tasks when compared to the (SQ) scholastic training. In other words, did pre to posttest WM progress depend on training type? There was a statistically significant interaction effect of time by type of training for both composite simple and complex span (see Table 6). Post hoc Tukey HSD tests revealed that these effects were significant in both WM

Table 6. Summary of main effects and interactions for composite WM test measures

Composite simple span				
Main effects	*F*	*df*	*p*	η²
Type of training	2.71	1,84	.10	0.03
Cognitive group	79.82	1,84	< **.0001**	0.48
Time	75.22	1,84	< **.0001**	0.47
Interactions				
Time*type of training	10.61	1,84	**.002**	0.11
Time*cognitive group	2.12	1,84	.14	0.02
Time*type of training*cognitive group	0.10	1,84	.32	0.01
Composite complex span				
Main effects	*F*	*df*	*p*	η²
Type of training	1.62	1,84	.21	0.02
Cognitive group	50.81	1,84	< **.0001**	0.38
Time	23.85	1,84	< **.0001**	0.22
Interactions				
Time*type of training	12.44	1,84	**.001**	0.13
Time*cognitive group	0.05	1,84	.83	0.00
Time*type of training*cognitive group	0.15	1,84	.70	0.00

training groups, with composite posttest scores being significantly higher than composite pretest scores, as seen in Table 8. In the SQ training group, there was a significant difference between pre and posttest scores for the TD children for simple span, but not for complex span. While pre to posttest WM gains are unexpected for TD_{SQ} children, it should be noted that the improvement for this particular group of participants is less statistically prominent than it is for DLD/TD_{WM} children when the *p*-value and Cohen's *d* are considered. No significant differences were found for DLD_{SQ} children.

Analyzing WM tests individually, a significant time by type of training interaction effect was observed for serial order word span, forward and backward digit spans (see Table 7). For counting span, this interaction was tendential and for non-word repetition it was not significant. Tukey HSD results confirmed that it was only the WM training groups that made significant pre to posttest progress on these individual WM measures, with DLD_{WM} children demonstrating posttest performance that was significantly better for serial order word span, forward digit recall, nonword repetition, and backward digit recall. For TD_{WM} children, there was significant improvement in serial order word span[11] and forward digit recall. No significant progress was observed for any of the WM tests for the DLD/TD_{SQ} children. These results are summarized in Appendix F.

Table 7. Summary of main effects and interactions for the individual WM test measures

Serial-order word span				
Main effects	F	df	p	η²
Type of training	7.94	1,84	**.006**	0.09
Cognitive group	47.99	1,84	**< .0001**	0.36
Time	34.90	1,84	**< .0001**	0.29
Interactions				
Time*type of training	4.53	1,84	**.04**	0.05
Time*cognitive group	4.98	1,84	**.03**	0.06
Time*type of training*cognitive group	0.10	1,84	.75	0.00
Forward digit span				
Main effects	F	df	p	η²
Type of training	2.80	1,84	.10	0.03
Cognitive group	64.49	1,84	**< .0001**	0.43
Time	31.68	1,84	**< .0001**	0.27
Interactions				
Time*type of training	5.16	1,84	**.03**	0.06
Time*cognitive group	1.17	1,84	.28	0.01
Time*type of training*cognitive group	0.00	1,84	.97	0.00
Nonword repetition				
Main effects	F	df	p	η²
Type of training	0.47	1,84	.49	0.01
Cognitive group	53.34	1,84	**< .0001**	0.39
Time	12.86	1,84	**.001**	0.13
Interactions				
Time*type of training	1.78	1,84	.19	0.02
Time*cognitive group	1.05	1,84	.31	0.01
Time*type of training*cognitive group	1.78	1,84	.19	0.02
Backward digit span				
Main effects	F	df	p	η²
Type of training	3.05	1,84	.08	0.04
Cognitive group	28.72	1,84	**< .0001**	0.25
Time	6.75	1,84	**.01**	0.07
Interactions				
Time*type of training	12.61	1,84	**.001**	0.13

(*Continued*)

Table 7. (*Continued*)

Backward digit span				
Main effects	F	df	p	η²
Time*cognitive group	0.39	1,84	.54	0.00
Time*type of training*cognitive group	0.57	1,84	.45	0.01
Counting span				
Main effects	F	df	p	η²
Type of training	0.27	1,84	.61	0.00
Cognitive group	34.08	1,84	< **.0001**	0.29
Time	10.62	1,84	**.002**	0.11
Interactions				
Time*type of training	3.26	1,84	.07	0.04
Time*cognitive group	1.69	1,84	.20	0.02
Time*type of training*cognitive group	0.03	1,84	.87	0.00

Table 8. Summary of pretest vs. posttest comparisons for the composite scores for simple and complex spans

		DLD$_{WM}$	DLD$_{SQ}$	TD$_{WM}$	TD$_{SQ}$
		M (%)(SD)	M (%)(SD)	M (%)(SD)	M (%)(SD)
Simple composite span	**PRE**	37.92 (9.21)	37.52 (8.84)	57.33 (12.89)	54.96 (10.65)
	POST	45.65 (9.43)	39.52 (10.34)	65.68 (12.56)	60.27 (10.10)
	Tukey HSD	***d* = 0.83**	ns	****d* = 0.66**	**d* = 0.51**
Complex composite span	**PRE**	23.34 (7.40)	21.72 (10.26)	39.79 (13.81)	41.16 (16.31)
	POST	30.21 (9.98)	23.06 (9.10)	47.74 (16.35)	42.21 (16.59)
	Tukey HSD	****d* = 0.78**	ns	***d* = 0.52**	ns

*** $p < .001$; ** $p < .01$; * $p < .05$

The repeated measures ANOVAs were also used to provide insight about a potential two-way interaction between time (pretest/posttest) and cognitive group (DLD/TD) and a potential three-way interaction between time, cognitive group, and training type (WM/SQ), see Tables 6 and 7. With the exception of a significant time by cognitive group interaction for the serial order word span measure (TD group improving most), these analyses revealed no significant interaction effects. These
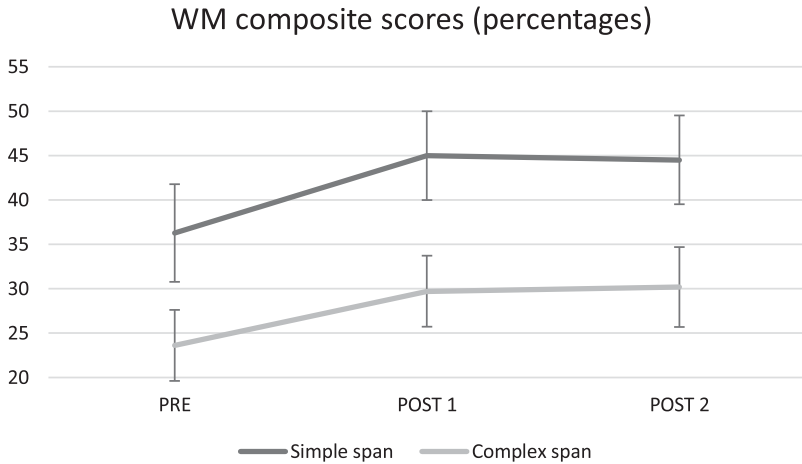
WM composite scores (percentages)

Figure 3. Distribution of composite WM scores for the pretests, immediate posttests (POST 1), and delayed posttests (POST 2).

results are understandable because pre to posttest improvements on the WM measures were not limited to DLD_{WM} children and were also observed in TD children.

Direct effects: Did the training results persist beyond immediate posttests?

To investigate if the observed gains in WM capacity persisted beyond the immediate posttests, paired-sample t-tests were run comparing the immediate posttest scores to those from the delayed posttests for twelve[12] of the *Magic Memory* participants. Before comparing the immediate posttest scores to the delayed posttest scores for these participants, we first wanted to verify that the observed pre to posttest gains still held for this subset of participants. The results confirm that for composite simple span: posttest performance was better than pretest performance even if the difference is only tendential, $t(22) = 2.08$, $p = .05$, $d = 0.85$. For composite complex span, a significant difference between pre and posttest performance was not observed, although there was a slight tendency, despite the small sample size, for the children in this subgroup to have higher scores on the posttests than on the pretests, $t(22) = 1.80$, $p = .08$, $d = 0.74$. Next, we compared immediate posttest scores to those of the delayed posttests, and our results showed that these differences were not significant ($p = .91$ for simple span and $p = .90$ for complex span), indicating that the WM gains seem to be relatively maintained over time, which is illustrated in Figure 3.

Transfer effects: Is there an improvement in syntactic tasks?

Analyses were then performed to examine whether training effects were limited to the domain of WM, or whether there was generalization to less directly related tasks, i.e., syntax, which was the primary aim of our study. For the sentence repetition task, the measures that were of interest were the percentage of correctly repeated

Table 9. Summary of main effects and interactions for the syntax measures

Correctly repeated syllables				
Main effects	F	df	p	η²
Type of training	0.62	1,84	.43	0.01
Cognitive group	94.64	1,84	< **.0001**	0.53
Time	5.54	1,84	**.02**	0.06
Interactions				
Time*type of training	9.68	1,84	**.003**	0.10
Time*cognitive group	0.00	1,84	.97	0.00
Time*type of training*cognitive group	0.32	1,84	.57	0.00
Respected target structure				
Main effects	F	df	p	η²
Type of training	0.28	1,84	.60	0.00
Cognitive group	130.86	1,84	< **.0001**	0.61
Time	12.66	1,84	**.001**	0.13
Interactions				
Time*type of training	3.43	1,84	.07	0.04
Time*cognitive group	0.51	1,84	.48	0.01
Time*type of training*cognitive group	0.11	1,84	.74	0.00
Respected degree of embedding				
Main effects	F	df	p	η²
Type of training	0.08	1,84	.77	0.00
Cognitive group	140.94	1,84	< **.0001**	0.63
Time	15.16	1,84	**.0002**	0.15
Interactions				
Time*type of training	2.70	1,84	.10	0.03
Time*cognitive group	0.63	1,84	.43	0.01
Time*type of training*cognitive group	0.01	1,84	.91	0.00

syllables, the percentage of respected target structures (subject and object relatives combined), and the percentage of structures in which the degree of embedding was respected (one, two, and three degrees of embedding combined). A repeated measures ANOVA showed significant main effects of cognitive group (TD > DLD) and time (posttest > pretest) for all three of these measures, but no significant main effect of type of training (Table 9).

Post hoc Tukey HSD tests, summarized in Table 10, showed that only DLD$_{WM}$ children significantly improved on these syntactic measures: for syllables, for

Table 10. Summary of pretest vs. posttest comparisons for complex sentence repetition

		DLD$_{WM}$	DLD$_{SQ}$	TD$_{WM}$	TD$_{SQ}$
		M (%)(SD)	M (%)(SD)	M (%)(SD)	M (%)(SD)
Percentage correct syllables	PRE	66.67 (15.50)	64.05 (16.24)	90.21 (10.43)	93.76 (3.83)
	POST	71.71 (15.13)	62.90 (18.29)	94.37 (4.97)	93.62 (4.03)
	Tukey HSD	*$d = 0.36$	ns	ns	ns
Percentage correct structures	PRE	35.21 (25.07)	29.00 (27.38)	80.37 (21.63)	88.15 (11.56)
	POST	45.00 (22.91)	33.67 (28.49)	88.89 (12.10)	89.26 (8.29)
	Tukey HSD	*$d = 0.42$	ns	ns	ns
Percentage correct degree of embedding	PRE	28.33 (20.18)	24.00 (24.70)	75.19 (22.96)	81.48 (12.64)
	POST	36.88 (22.78)	28.33 (25.21)	81.85 (17.54)	83.33 (10.79)
	TukeyHSD	*$d = 0.42$	ns	ns	ns

*$p < .05$

respected structures, and for embedding. There was no significant syntactic improvement for TD$_{WM}$ children or for the TD/DLD$_{SQ}$ children.

The results presented in Table 9 also showed a significant time by type of training interaction effect for the percentage of correctly repeated syllables, but as previously stated, post hoc Tukey HSD tests revealed that significant improvement occurred only in the DLD$_{WM}$ group. This interaction effect is likely explained by the significant main effect of time found in the children with DLD$_{WM}$ and by the more modest but consistent improvement of the TD$_{WM}$ children. For the other two variables of interest (percentage of respected structures and embedding), no significant time by type of training interaction effects were observed. There were also no significant time by cognitive group or time by type of training by cognitive group interactions.

Transfer effects: Did the training results persist beyond immediate posttests?

As with the WM tests, we wanted to investigate whether gains observed in the repetition task endured from the immediate posttest phase to the delayed posttest phase. To do this, we first wanted to ensure that the previously observed pre to immediate posttest gains were still present in our reduced sample of twelve participants, but paired-sample *t*-tests revealed that this was not the case ($p = .55$ for syllables, $p = .49$ for structure, and $p = .28$ for embedding). Consequently, in this subgroup of participants, we were unable to make a meaningful comparison between the immediate posttests and the delayed posttests for the repetition task.

Table 11. Summary of comparisons made for pretest measures of differing syntactic complexity when the WM and scholastic training participants are put together in each cognitive group

	DLD		TD	
	M (%)(SD)	comparison	*M (%)(SD)*	comparison
−RC	31.25 29.03)	***d = 1.47**	86.11 (23.68)	ns
+RC	10.51 (14.54)		91.98 (7.95)	
Subject relatives	45.00 (35.40)	***d = 0.64**	85.56 (17.64)	ns
Object relatives	25.77 (23.21)		83.61(20.45)	
1 degree of embedding	35.54 (28.72)	***d = 1.06**	90.12 (17.77)	***d = 1.31**
2–3 degrees of embedding	11.86 (15.94)		60.65 (26.47)	

***$p < .001$

Transfer effects: Are the effects modulated by syntactic complexity factors?

While our main results confirmed that our WM training regime was successful when simple and complex structures (i.e., subject and object relatives) were combined, the goal of our final analyses was to distinguish between more or less complex sentences within our task and to examine if training effects were different with respect to this distinction. To recall, we predicted that these effects would be less obvious for simpler syntactic structures that are hypothesized to be less taxing for WM. To investigate this, all of the items in the repetition task were split into groups by type of sentence (simple sentences containing no relative clause vs. complex sentences containing a relative clause), and complex sentences were split into groups by type of structure (subject vs. object) and degree of embedding (one degree vs. two-three degrees). Before analyzing pre to posttest performance in our participants, we first wanted to verify that complexity effects were in fact present in pretest performance, see Table 11. Paired-sample *t*-tests revealed that in the pretest phase, children with DLD correctly repeated significantly more syllables in simple sentences without a relative clause (−RC) than in ones containing a relative clause (+RC) which was not the case for TD children who were at ceiling for both measures. Similarly, for what concerns target structure, a complexity effect was only found for children with DLD, who correctly repeated significantly more subject relatives than object relatives in the pretest phase. For TD children, there was no significant difference between these two measures prior to training. However, for degree of embedding, both TD and DLD groups produced significantly more sentences containing one degree of embedding than those containing either two or three degrees in the pretest phase.

Table 12. Summary of additional pretest vs. posttest comparisons on measures of differing degrees of syntactic complexity

Percentage correctly repeated sentences		DLD$_{WM}$	DLD$_{SQ}$	TD$_{WM}$	TD$_{SQ}$
		M (%)(SD)	M (%)(SD)	M (%)(SD)	M (%)(SD)
-RC	PRE	34.38 (30.62)	26.25 (26.25)	82.64 (30.36)	89.58 (14.38)
	POST	35.16 (30.70)	32.50 (30.99)	84.72 (24.46)	87.50 (15.46)
	Tukey HSD	ns	ns	ns	ns
Subject relatives	PRE	50.00 (32.53)	37.00 (39.08)	82.22 (21.57)	88.89 (12.31)
	POST	60.62 (30.68)	44.00 (37.05)	92.22 (10.03)	91.11 (12.32)
	Tukey HSD	ns	ns	ns	ns
Object relatives	PRE	26.25 (22.25)	25.00 (25.24)	79.44 (24.37)	87.78 (15.17)
	POST	37.18 (24.79)	27.50 (26.73)	86.67 (17.82)	88.33 (10.43)
	Tukey HSD	**$d = 0.47$	ns	ns	ns
1 degree of embedding	PRE	39.24 (25.55)	32.22 (33.41)	88.89 (21.89)	91.36 (12.96)
	POST	50.35 (27.36)	37.22 (31.89)	93.21 (12.72)	96.30 (6.60)
	Tukey HSD	*$d = 0.42$	ns	ns	ns
2–3 degrees of embedding	PRE	11.98 (17.06)	11.67 (14.41)	54.63 (31.21)	66.67 (19.80)
	POST	16.67 (21.17)	14.17 (20.43)	64.81 (27.94)	63.89 (22.32)
	Tukey HSD	ns	ns	ns	ns

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

Next, we investigated if improvement in the repetition task in the posttest phase was related to degree of complexity of the sentence. We have already reported a significant main effect of time for sentences containing a relative clause in DLD children, but repeated measures ANOVA showed that no significant progress was made for simple sentences without a relative clause (−RC), $F(1,84) = 0.58$, $p = .45$, $\eta^2 = 0.01$. In other words, significant improvement only occurred for syntactically complex sentences either containing a subject or an object relative. As for different types of relatives, we found a significant main effect of time for the repetition of subject relatives, $F(1,84) = 5.56$, $p = .02$, $\eta^2 = 0.06$, but with post hoc Tukey HSD analysis it was not possible to statistically distinguish pre and posttest performance in any of

the groups, see Table 12. When the same analyses were run for pre and posttest performance of object relatives, a significant main effect of time was also observed $F(1,84) = 9.78$, $p = .002$, $\eta^2 = 0.10$, as well as a significant time by type of training interaction, $F(1,84) = 4.96$, $p = .03$, $\eta^2 = 0.06$. Post hoc Tukey HSD tests (Table 12) revealed a significant pre to posttest difference in only DLD$_{WM}$ children. Finally, for degree of embedding, there was a significant main effect of time for sentences containing a relative clause and one degree of embedding, $F(1,84) = 11.76$, $p = .001$, $\eta^2 = 0.12$, and post hoc Tukey HSD tests confirmed that the pre to posttest improvement was significant only for DLD$_{WM}$ children. For items containing two-three degrees of embedding, no significant effects were observed. For the most part, these results are in line with our predictions, with the exception of degree of embedding. For DLD$_{WM}$ children, significant improvement was observed for more complex sentences containing a relative clause but not for less complex ones without a relative clause, and for more complex object relatives but not for less complex subject relatives.

Discussion

Our study explored the effects on a new intensive WM training, *Magic Memory*, on WM (assessed via untrained tasks) and mastery of complex syntax (assessed via repetition of relative clauses). This training was administered to 18 TD children aged 6 to 12 years and to 32 children with DLD of the same age. Strong predictive links between WM and syntax have been repeatedly demonstrated, for TD children (Delage & Frauenfelder, 2019; Marinis & Saddy, 2013; Montgomery et al., 2008) and also for children with DLD (Durrleman & Delage, 2016; Frizelle & Fletcher, 2015; Zebib et al., 2019). In the latter case, predictive links between WM and complex syntax (Delage & Frauenfelder, 2020) may be suggestive of novel therapeutic avenues, in which case we expect improvements in memory abilities to minimize the persistent syntactic deficits associated with DLD. Finally, we compared the effect of WM training to that of age-matched children (18 TD, 20 DLD) who followed an alternative, scholastic training, for whom no syntactic improvement was expected, which would allow us to conclude that the potential positive effects observed in WM trained children stemmed specifically from our WM program. Our original, computerized intervention targeted the aspects of WM shown to predict complex syntax processing in both TD and DLD children, namely serial order memory and dual-task processing (Delage & Frauenfelder, 2019, 2020).

Direct effects

Our results confirm our main predictions, specifically that *Magic Memory* improves WM performance in children with TD and DLD and leads to better repetition of complex sentences in children with DLD. More precisely, a positive direct effect of this program on WM performance was observed for untrained tasks in both TD$_{WM}$ and DLD$_{WM}$ children. Furthermore, a time by training group interaction was found for both simple and complex-span composite scores, which shows that the observed effects are specific to the training program and cannot be attributed to external factors such as maturation or motivation. It is nevertheless true that the WM tasks for

which we showed direct effects resemble the activities trained by the WM program, although the visual and verbal material differ, as well as the test modality (paper versus computer-based). However, it is important to note that such direct effects have not always been found in the literature (see Melby-Lervag & Hulme, 2013; Melby-Lervag et al., 2016), and it is thus noteworthy that children seem to have progressed sufficiently to transfer skills from one support to another. That previous studies did not detect the benefits observed here could be due to the fact that these studies have been mainly conducted with healthy adults and moreover involved generic WM training programs with many visuospatial activities, such as is the case for *Jungle Memory* and *Cogmed*. Put differently, our more convincing results may stem from our program's focus on the specific underlying components of WM found to establish a predictive relationship with syntax in children with difficulties both in WM and syntax. That we do not find an interaction between training type and the cognitive group (TD/DLD) suggests that the two groups respond similarly to training. However, an unexpected result is that TD_{SQ} children also progressed for the simple composite span, which was not observed in DLD_{SQ} children who followed the same alternative training. This progression in TD children could be explained by their normal maturation of simple spans. However, when considering progression in each WM task separately, no significant progression is observed in children who followed the scholastic training, whereas children with DLD_{WM} showed such a progression for 4/5 tasks and TD_{WM} progressed on 2/5 tasks. Finally, we were also able to test for long-term effects of the WM training in evaluating WM performance three months after the end of the training. Even though we could test only 12 of the 32 DLD_{WM} children, these delayed posttests showed that gains in WM seem to be maintained. These results are encouraging regarding the program's long-term effectiveness, particularly since such effects are generally absent from the literature (Melby-Lervag & Hulme, 2013). However, analyses conducted with a larger cohort of participants that compare their long-term performance with that of control participants are needed to validate these initial observations.

Transfer effects

As for indirect effects on syntactic capacities, we showed a significant effect of time for the measures analyzed in sentence repetition, with posttest performance being better than pretest performance in children with DLD_{WM}, whereas no improvement was observed in children following the alternative training. Amongst these measures, a progression in the percentage of correctly repeated syllables can be explained by the progression of WM abilities, particularly the phonological loop, as these children managed to repeat more digits or nonwords in posttest than pretest. However, DLD_{WM} children also improved their capacity to reproduce the target structure and the expected level of embedding (without considering lexical substitutions or omissions), which is clearly less related to the simple maintenance and recall of the phonological loop, and more to the information processing required for complex WM tasks. Furthermore, that our transfer effects show a progression on sentence repetition cannot be attributed to similarities between the material of *Magic Memory* and this syntactic test, as may be claimed to be the case for WM tests, since *Magic*

Memory only includes isolated words as verbal material. It should be noted, however, that despite the progress of children with DLD_{WM}, they still perform at a much lower level than TD children at posttests, with scores of 45% for the target structures and 37 % for the degree of embedding, compared to 80% and above for TD children. As for TD children, we were not able to show significant transfer effects in the group who followed the WM training, probably because their performance level was already high in the pretest phase. However, close inspection of Table 8 reveals that this TD_{WM} group still progressed more than the TD_{SQ} children. The fact that numbers of participants vary between TD (18_{WM}, 18_{SQ}) and DLD (32_{WM}, 20_{SQ}) groups can help to explain why more differences are significant in the DLD group than the TD one, although there may be more to the story. Indeed, the finding that transfer effects are only significant in children with DLD and not in TD is consistent with the literature, which has tended to report the absence of such effects in participants without language impairment (Melby-Lervag et al., 2016). It would seem then that WM training might increase syntactic abilities most when there is a specific deficit in this realm.

Modulation of training effects

We predicted that the more complex sentences were particularly taxing for WM and thus should be the ones that improve the most through WM training. In line with this prediction, we found a significant progression in DLD_{WM} children in repetition of complex sentences, while such a progression is absent for the repetition of simple sentences (without embedding), which were crucially of the same length as the complex ones. Moreover, we found an interaction of time by type of training for object relatives and not for subject relatives. It thus seems that more complex sentences, with a disruption of the canonical word order as it is the case for object relatives, may be more sensitive to WM training than simpler structures such as subject relatives. Here again, this effect is essentially due to DLD_{WM} children, the only group to demonstrate a significant progression for object relatives between pre- and posttest. Note that we consider the disruption of the canonical order attested in object relative clauses to be a factor of complexity, in line with the DCH approach sketched by Jakubowicz (2005, 2011). However, as mentioned by an anonymous reviewer, the complexity of object relatives can also be linked to the intervention of a similar argument, along the lines of Featural Relativized Minimality (fRM, Rizzi, 2004; Friedman et al., 2009). Future studies should go one step further by examining the effect of our WM training on different types of object structures, namely those with and without syntactic intervention as described by fRM, to see if the training boosts all object structures similarly, or if the nature of syntactic intervention (i.e., featural similarity) modulates this effect.

As for the correct repetition of the target level of embedding, it is interesting that the DLD_{WM} group significantly progressed only for the sentences containing one level of embedding (39% at pretest, 50% at posttest), and not for additional levels of embedding which can be explained by floor effects on these measures (with percentages inferior to 17% for 2–3 levels of embedding). Although not significant, the pattern seems different in the TD_{WM} group: they were at ceiling at pretest for level 1 (89%), which leaves too little room for improvement, yet they showed an

improvement of 10% at posttest for sentences with 2–3 levels of embedding for which they were not initially at ceiling (55% at pretest, 65% at posttest). It is clear that our TD groups are too small in number for such differences to be significant, which is a limitation of our study. For future studies, it also seems important to test the effect of WM training on comprehension of such complex structures, since children with DLD also show difficulties in this realm (as in Delage & Frauenfelder, 2020 or Friedman & Novrogrodsky, 2004). The gains may be subtler in comprehension, as an asymmetry between comprehension and production has been reported in this direction for children with DLD (Jakubowicz & Tuller, 2008; Contemori & Garrafa, 2010). Still, if our WM training can improve performance in both modalities, i.e., by giving rise not only to gains in production but also to (possibly subtler) gains in comprehension, this would argue more forcefully for the role of executive functions on the language capacities in DLD.

Overall, our results are consistent with those obtained in previous studies, which highlighted the modulation between the degree of syntactic complexity and WM capacities. In 2019, Delage and Frauenfelder specifically demonstrated that WM evaluated in 48 TD children aged 6–12 years explained a greater part of the variance for relatives with higher levels of embedding (54%) than for relatives with lower level of embedding (38%). In the same vein, the production of subject relatives was not predicted by WM, whereas both simple and complex spans explained a significant part of the variance of results in repetition of object relative clauses. While this prior study was a cross-sectional one, the current study is the first, to our knowledge, to demonstrate this link using a longitudinal, training paradigm. One should ask, however, if a targeted training focusing specifically on complex syntax would produce the same effects. In that sense, it would be interesting to compare the effect of systematic training in WM with that of direct syntactic training (with, for example, the metalinguistic approach of Shape Coding, Ebbels, 2007). There is still little evidence in the scientific literature regarding the effectiveness of syntax-based speech-language pathology therapy (Ebbels, 2014; Law et al., 2003), but such trainings, including explicit grammatical interventions, seem to be efficient (Balthazar & Scott, 2018; Calder et al., 2020; Zwitserlood et al., 2015).

Long-term effects

Our findings also suggest potential long-term effects of WM training, given that direct effects seem to be maintained after three months, although this needs to be interpreted with extreme caution given that only 12 participants were retested with delayed posttests. If future studies confirm that children retain the benefits of the training and continue to progress in syntax, the training could be seen as a cognitive "boost" allowing the children to free up resources to process complex syntax. Future work should explore precisely what underlying cognitive processes are enhanced as a result of the training program. Indeed, WM is underpinned by a whole set of attentional and executive processes and more specifically by selective attention (Majerus et al, 2009; Veer et al., 2017). It is therefore possible that with a training program as intensive as ours, requiring sustained attention for 30 minutes per session, it is the attentional component, included in all of the WM models (Baddeley, 2003; Barrouillet et al., 2004; Cowan, 1999), that has been particularly

called upon. Although selective attention is less studied than WM in its relationship to language development, some studies have already shown a link between attentional abilities and relative sentence comprehension in preschool children (Finney et al., 2014). With respect to children with DLD, it is known that a significant proportion of these children suffer from attentional difficulties (Ebert & Kohnert, 2011), suggesting that there is a significant interaction between language and attention disorders. It would therefore be interesting to use pure training of the attentional component as an alternative training, which would make it possible to determine more precisely whether the underlying component necessary for this cognitive boost is indeed the attentional component or, on the contrary, whether it is not sufficient to improve language. Additionally, it seems important to include measures of attention in future work in order to understand how intersubject variability in this domain can affect the progression of trained children. This is also the case for other co-occurring problems frequently found in children with DLD such as speech sound disorders (Pennington & Bishop, 2009).

Clinical applications

Finally, this work presents potential clinical implications. Indeed, if future work can replicate our results, and notably can demonstrate robust long-term effects, the logical next step would be to make the training material available to clinicians, which would provide them with innovative and scientifically validated material for the therapy of children with syntactic disorders. The idea here would not be to replace conventional therapy with a computer-based WM protocol, but rather to use it as a complement to this treatment, in the vein of an Evidence-Based Practice approach, consisting of applying the results of research to clinical decision making (Sackett et al., 2000).

Notes

1. We consider (verbal) WM as referring to both the phonological loop (assessed by simple spans) and the central executive (assessed by complex spans), in line with the work of Barouillet and Camos (2007).
2. Derivational Complexity Metric (Jakubowicz, 2011):
   a. Merging α_i n times gives rise to a less complex derivation than merging α_i $(n + 1)$ times.
   b. Internal merge of α gives rise to a less complex derivation than internal merge of $\alpha + \beta$.
3. Such as "Show me the lady$_i$ that the girl is kissing t$_i$", as compared to "Show me who$_i$ the girl is kissing t$_i$".
4. We chose to use a sentence repetition task because it allows us to precisely evaluate the ability of children to produce structures in which these two factors are manipulated, which is not the case for spontaneous language analysis or elicitation tasks. Hence, in spontaneous language, children have the "choice" of the structures used. If, for example, they produce no object relative clauses in their corpora, this does not mean

that they are unable to produce such sentences. As for elicited production, it seems impossible to elicit relative clauses with two or three levels of embedding. Indeed, the previous studies only elicit subject and object relatives with only one level of embedding (see, for example, Friedman et al., 2009).

5. Note that finding participants with DLD who met this criterion was not particularly challenging. Only one child with DLD was formally excluded from participation in this study as his WM scores on the inclusion tests were age appropriate despite confirmed impairment in syntax.

6. Only one TD child was excluded from the study due to low WM skills. This child had a family history of DLD. Knowing the genetic factors contributing to DLD (see Bishop, 2006), we suspected that this child may have undiagnosed DLD.

7. Note, however, that WM skills of DLD_{SQ} children tended to be lower than those of DLD_{WM}. This difference is no longer present in the comparison of pretests evaluating the same components, as explained further.

8. Additional sentences were added to the original task to provide the two matched versions.

9. These nine sentences were divided into three so-called "0-level relatives", three "pseudo-relatives", and three "genuine relatives". First, the so-called "0-level relatives" are the least complex structures since they are not embedded within a matrix clause. Next, pseudo-relatives, consisting of presentational structures with *y'a* and *c'est,* have an intermediate status between 0-level relatives and genuine relatives. Indeed, since these involve embedding of a subordinate clause within a DP, which is itself inside an IP, pseudo-relatives involve a flatter structure, i.e., less deep embedding (see Delage et al., 2008 and Hamann & Tuller, 2014 for detailed analyses of such structures).

10. An anonymous reviewer suggested that the analyses for the direct and transfer effects should perhaps be conducted on the monolingual and bilingual participants separately. However, as we verified that all of the main effects remained the same when the bilingual participants were removed from the sample, the results we present here include both monolingual and bilingual participants.

11. The good progress made by the TD children in serial order word span is also found in the comparison of their scores at posttests. While the two TD groups did not differ at pretest for this measure, $p = .16$, they differed significantly at posttest, with better scores for children in the WM group, $t(34) = -2.1, p = .04$.

12. Due to various familial constraints and the exploratory nature of this analysis, only 12 of the 32 DLD participants in the WM training group were included in the delayed posttests.

References

Adams, A. M., & Gathercole, S. E. (2000). Limitations in Working Memory: Implications for Language Development. *International Journal of Language and Communication Disorders*, **35**(1), 95–117.

Adani, F., Stegenwallner-Schütz, M., Haendler, Y., & Zukowski, A. (2016). Elicited production of relative clauses in German: Evidence from typically developing children and children with specific language impairment. *First Language*, **36**(3), 203–227.

Alloway, T. P., & Gathercole, S. (2005). Working memory and short-term sentence recall in young children. *European Journal of Cognitive Psychology*, **17**, 207–220.

Archibald, L. M., & Gathercole, S. E. (2006). Short-term and working memory in specific language impairment. *International Journal of Language & Communication Disorders*, **41**(6), 675–693.

Archibald, L. M., & Harder Griebeling, K. (2016). Rethinking the connection between working memory and language impairment. *International Journal of Language & Communication Disorders*, **51**(3), 252–264.

Archibald, L. M., & Joanisse, M. F. (2009). On the sensitivity and specificity of nonword repetition and sentence recall to language and memory impairments in children. *Journal of Speech, Language, and Hearing Research*, **52**, 899–914.

Armon-Lotem, S., & Meir, N. (2016). Diagnostic accuracy of repetition tasks for the identification of specific language impairment (SLI) in bilingual children: Evidence from Russian and Hebrew. *International Journal of Language & Communication Disorders*, **51**, 715–731.

Arosio, F., Guasti, M. T., & Stucchi, N. (2011). Disambiguating information and memory resources in children's processing of Italian relative clauses. *Journal of Psycholinguistic Research*, **40**(2), 137–154.

Baddeley, A. D. (2003). Working memory and language: an overview. *Journal of Communication Disorders*, **36**(3), 189–208.

Baird, G., Dworzynski, K., Slonims, V., & Simonoff, E. (2010). Memory impairment in children with language impairment. *Developmental Medicine & Child Neurology*, **52**(6), 535–540.

Balthazar, C. H., & Scott, C. M. (2018). Targeting complex sentences in older school children with specific language impairment: Results from an early-phase treatment study. *Journal of Speech, Language, and Hearing Research*, **61**(3), 713–728.

Barouillet, P., & Camos, V. (2007). Le développement de la mémoire de travail. In J. Lautrey (Ed.), Psychologie du développement et de l'éducation (pp. 51–86). Paris: Presses Universitaires.

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, **133**(1), 83.

Bentea, A., Durrleman, S., & Rizzi, L. (2016). Refining intervention: The acquisition of featural relations in object A-bar dependencies. *Lingua*, **169**, 21–41.

Bishop, D. V. (2006). What causes specific language impairment in children? *Current Directions in Psychological Science*, **15**(5), 217–221.

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

Boutard, C., & Gatignol, P. (2015). *Mémoire de Travail*. Ortho Editions.

Calder, S. D., Claessen, M., Ebbels, S., & Leitão, S. (2020). Explicit grammar intervention in young school-aged children with developmental language disorder: An efficacy study using single-case experimental design. *Language, Speech, and Hearing Services in Schools*, **51**(2), 298–316.

Case, R., Kurland, M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory. *Journal of Experimental Child Psychology*, **33**, 386–404.

Chomsky, N. (2005). Three factors in language design. *Linguistic Inquiry*, **36**(1), 1–22.

Contemori, C., & Garrafa, M. (2010). Comparison of modalities in SLI syntax: A study on the comprehension and production of non-canonical sentences. *Lingua*, **120**(8), 1940–1955.

Conti-Ramsden, G., & Durkin, K. (2008). Language and independence in adolescents with and without a history of specific language impairment (SLI). *Journal of Speech, Language, and Hearing Research*, **51**(1), 70–83.

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

Cowan, N. (1999). An Embedded-Processes Model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). Cambridge University Press.

De Abreu, P. M. J. E., Gathercole, S. E., & Martin, R. (2011). Disentangling the relationship between working memory and language: The roles of short-term storage and cognitive control. *Learning and Individual Differences*, **21**(5), 569–574.

Delage, H., Da Costa, J., & Durrleman, S. (2017). Magic Memory. *Experimental program.* http://tecfa.unige.ch/proj/logo/MagicMemory.html

Delage, H., & Frauenfelder, U. H. (2019). Syntax and working memory in typically-developing children: Focus on syntactic complexity. *Language, Interaction and Acquisition*, **10**(2), 141–176.

Delage, H., & Frauenfelder, U. H. (2020). Relationship between working memory and complex syntax in children with Developmental Language Disorder. *Journal of Child Language*, **47**(3), 600–632.

Delage, H., Monjauze, C., Hamann, C., & Tuller, L. (2008). Relative Clauses in Atypical Acquisition of French. *Language Acquisition and Development: Proceedings of GALA 2007* (pp. 166–176). Cambridge Scholars Publishing.

Durrleman, S., & Delage, H. (2016). Autism Spectrum Disorder and Specific Language Impairment: Overlaps in syntactic profiles. *Language Acquisition*, **23**, 361–386.

Durrleman, S., Delage, H., Prévost, P., & Tuller, L. (2017). The comprehension of passives in autism spectrum disorder. *Glossa*, **2**, 1–30.

Ebbels, S. (2007). Teaching grammar to school-aged children with specific language impairment using shape coding. *Child Language Teaching and Therapy*, **23**(1), 67–93.

Ebbels, S. (2014). Effectiveness of intervention for grammar in school-aged children with primary language impairments: A review of the evidence. *Child Language Teaching and Therapy*, **30**(1), 7–40.

Ebert, K. D., & Kohnert, K. (2009). Non-linguistic cognitive treatment for primary language impairment. *Clinical Linguistics & Phonetics*, **23**(9), 647–664.

Ebert, K. D., & Kohnert, K. (2011). Sustained attention in children with primary language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, **54**(5), 1372–1384.

Evans, J. L., & Brown, T. T. (2016). Specific language impairment. In S. Small & G. Hickok (Eds.), *Neurobiology of language* (pp. 899–912). Academic Press.

Everitt, A., Hannaford, P., & Conti-Ramsden, G. (2013). Markers for persistent specific expressive language delay in 3–4-year-olds. *International Journal of Language & Communication Disorders*, **48**(5), 534–553.

Finney, M. C., Montgomery, J. W., Gillam, R. B., & Evans, J. L. (2014). Role of working memory storage and attention focus switching in children's comprehension of spoken object relative sentences. *Child Development Research*, **2014**.

Friedmann, N., Belletti, A., & Rizzi, L. (2009). Relativized relatives: Types of intervention in the acquisition of A-bar dependencies. *Lingua*, **119**(1), 67–88.

Friedmann, N., & Novogrodsky, R. (2004). The acquisition of relative clause comprehension in Hebrew: A study of SLI and normal development. *Journal of Child Language*, **31**, 661–681.

Friedmann, N., & Novogrodsky, R. (2008). Subtypes of SLI: SySLI, PhoSLI, LeSLI, and PraSLI. In A. Gavarró, & M. João Freitas (Eds.), *Language acquisition and development* (pp. 205–217). Cambridge Scholars Press.

Friedmann, N., & Novogrodsky, R. (2011). Which questions are most difficult to understand? The comprehension of Wh questions in three subtypes of SLI. *Lingua*, **121**(3), 367–382.

Frizelle, P., & Fletcher, P. (2015). The role of memory in processing relative clauses in children with specific language impairment. *American Journal of Speech and Language Pathology*, **24**, 47–59.

Gathercole, S. E., & Baddeley, A. (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language*, **29**, 336–360.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, **68**(1), 1–76.

Gillam, R. B., & Hoffman, L. M. (2003). Information processing in children with specific language impairment. In L. Verhoeven & H. van Bolkom (Eds.), *Classification of developmental language disorders* (pp. 149–170). Psychology Press.

Hamann, C., & Tuller, L. (2014). Genuine versus superficial relatives in French: the depth of embedding factor. *Research in Generative Grammar*, **36**, 47–82.

Henry, L. A., & Botting, N. (2017). Working memory and developmental language impairments. *Child Language Teaching and Therapy*, **33**(1), 19–32.

Hick, R. F., Botting N., & Conti-Ramsden G. (2005). Short-term memory and vocabulary development in children with Down syndrome and children with specific language impairment. *Developmental Medicine and Child Neurology*, **47**, 532–538.

Hoffman, L. M., & Gillam, R. B. (2004). Verbal and Spatial Information Processing Constraints in Children with and without Specific Language Impairment. *Journal of Speech, Language and Hearing Research*, **47**, 114–125.

Holmes, J., Butterfield, S., Cormack, F., van Loenhoud, A., Ruggero, L., Kashikar, L., & Gathercole, S. (2015). Improving working memory in children with low language abilities. *Frontiers in Psychology*, **6**, 519.

Jakubowicz, C. (2005). The Language Faculty: (ab)normal development and interface constraints. Talk presented at GALA, University of Siena.

Jakubowicz, C. (2011). Measuring derivational complexity: New evidence from typically developing and SLI learners of L1 French. *Lingua*, **121**(3), 339–351.

Jakubowicz, C., & Strik, N. (2008). Scope-marking strategies in the acquisition of long distance wh-questions in French and Dutch. *Language and Speech*, **51**(1–2), 101–132.

Jakubowicz, C., & Tuller, L. (2008). Specific Language Impairment in French. In D. Ayoun, (Ed.), *Studies in French Applied Linguistics* (pp. 97–134). John Benjamins.

Kail, R. (1994). A method for studying the generalized slowing hypothesis in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, **37**(2), 418–421.

Kapa, L. L., & Plante, E. (2015). Executive function in SLI: Recent advances and future directions. *Current Developmental Disorders Reports*, **2**(3), 245–252.

Kapa, L. L., Plante, E., & Doubleday, K. (2017). Applying an integrative framework of executive function to preschoolers with specific language impairment. *Journal of Speech, Language, and Hearing Research*, **60**(8), 2170–2184.

Khomsi, A., Khomsi, J., Parbeau-Guéno, A., & Pasquet, F. (2007). *Bilan informatisé de langage oral au Cycle III et au collège (BILO3C)*. ECPA.

Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition*, **2**(1), 15–47.

Law, J., Garett, Z., & Nye, C. (2003). Speech and language therapy intervention for child. *Cochrane data base of systematic review, 3.*

Law, J., McKean, C., Murphy, C. A., & Thordardottir, E. (Eds.). (2019). *Managing children with developmental language disorder: theory and practice across Europe and beyond.* Routledge.

Leclercq, A. L., & Majerus, S. (2010). The relationship between serial order STM and vocabulary development: A longitudinal study. *Developmental Psychology*, **46**, 417–427.

Leonard, L. B. (2014). Specific language impairment across languages. *Child Development Perspectives*, **8**, 1–5.

Mackey, A., & Sachs, R. (2012). Older learners in SLA research: A first look at working memory, feedback, and L2 development. *Language Learning*, **62**(3), 704–740.

Majerus, S. (2008). La mémoire verbale à court terme : un simple produit des interactions entre systèmes langagiers, attentionnels et de traitement de l'ordre sériel?. *Psychologie française*, **53**(3), 327–341.

Majerus, S. (2016). Optimisation et rééducation de la mémoire de travail : une synthèse critique, *Approche Neuropsychologique des Apprentissages chez l'Enfant*, **141**, 167–174.

Majerus, S., Heiligenstein, L., Gautherot, N., Poncelet, M., & Van der Linden, M. (2009). Impact of auditory selective attention on verbal short-term memory and vocabulary development. *Journal of Experimental Child Psychology*, **103**(1), 66–86.

Majerus, S., Poncelet, M., Greffe, C., & Van der Linden, M. (2006). Relations between vocabulary development and verbal short-term memory: The relative importance of short-term memory for serial order and item information. *Journal of Experimental Child Psychology*, **93**, 95–119.

Marini, A., Eliseeva, N., & Fabbro, F. (2019). Impact of early second-language acquisition on the development of first language and verbal short-term and working memory. *International Journal of Bilingual Education and Bilingualism*, **22**(2), 165–176.

Marinis, T., & Saddy, D. (2013). Parsing the passive: comparing children with Specific Language Impairment to sequential bilingual children. *Language Acquisition*, **20**(2), 155–179.

Melby-Lervag, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, **49**(2), 270–291.

Melby-Lervag, M., Redick, T. S., & Hulme, C. (2016). Working memory training does not improve performance on measures of intelligence or other measures of "far transfer" evidence from a meta-analytic review. *Perspectives on Psychological Science*, **11**(4), 512–534.

Montgomery, J. (2000). Verbal working memory in sentence comprehension in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, **43**, 293–308.

Montgomery, J., Magimairaj, B., & O'Malley, M. (2008). Role of working memory in typically developing children's complex sentence comprehension. *Journal of Psycholinguistic Research*, **37**(5), 331–354.

Montgomery, J. W., & Evans, J. L. (2009). Complex Sentence Comprehension and Working Memory in Children with Specific Language Impairment. *Journal of Speech Language and Hearing Research*, **52**(2), 269–288.

Morrison, A. B., & Chein, J. M. (2011). Does working memory training work? The promise and challenges of enhancing cognition by training working memory. *Psychonomic Bulletin & Review*, **18**(1), 46–60.

Mousty, P., Leybaert, J., Alegria, J., Content, A., & Moraïs, J. (1994). *B.E.L.E.C. : Batterie d'évaluation du langage et de ses troubles.* Laboratoire de psychologie.

Nicolson, R. I., & Fawcett, A. J. (2007). Procedural learning difficulties: reuniting the developmental disorders? *Trends in Neurosciences*, **30**(4), 135–141.

Nippold, M. A., Mansfield, T. C., Billow, J. L., & Tomblin, J. B. (2009). Syntactic development in adolescents with a history of language impairments: A follow-up investigation. *American Journal of Speech-Language Pathology*, **18**(3), 241–251.

Peijnenborgh, J. C., Hurksc, P. M., Aldenkamp, A. P., & Vlesabd, J. (2015). Efficacy of working memory training in children and adolescents with learning disabilities: A review study. *Neuropsychological Rehabilitation*, **26**(5–6), 645–672.

Pennington, B. F., & Bishop, D. V. (2009). Relations among speech, language, and reading disorders. *Annual Review of Psychology*, **60**, 283–306.

Petruccelli, N., Bavin, E. L., & Bretherton, L. (2012). Children with specific language impairment and resolved late talkers: Working memory profiles at 5 years. *Journal of Speech, Language, and Hearing Research*, **55**(6), 1690–1703.

Polišenská, K., Chiat, S., & Roy, P. (2015). Sentence repetition: What does the task measure? *International Journal of Language & Communication Disorders*, **50**(1), 106–118.

Poll, G. H., Miller, C. A., Mainela-Arnold, E., Adams, K. D., Misra, M., & Park, J. S. (2013). Effects of children's working memory capacity and processing speed on their sentence imitation performance. *International Journal of Language & Communication Disorders*, **48**(3), 329–342.

Raven, J. C., & Court, J. H. (1998). *Raven's progressive matrices and vocabulary scales*. Oxford Pyschologists Press.

Riches, N. G., Loucas, T., Baird, G., Charman, T., & Simonoff, E. (2010). Sentence repetition in adolescents with specific language impairments and autism: An investigation of complex syntax. *International Journal of Language & Communication Disorders*, **45**(1), 47–60.

Rizzi, L. (2004). Locality and left periphery. In A. Belletti (Ed.), *Structures and beyond: The cartography of syntactic structures* (pp. 223–251). Oxford University Press.

Sackett, D. L., Straus, S. E., Richardson, W. S., Rosenberg, W., & Haynes, R. B. (2000). *Evidence-based medicine: How to practice and teach EBM* (2nd ed.). Churchill Livingstone.

Scheidnes, M., & Tuller, L. (2014). L2 children embed normally. *New directions in the acquisition of Romance languages: Selected proceedings of the Romance Turn V*, 261.

Stanford, E., & Delage, H. (2019). Complex syntax and working memory in children with specific learning difficulties. *First Language, Special Issue: Syntax and Verbal Short-Term Memory in Developmental Disorders*, 1–26.

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.

Sung, J. E., Yoo, J. K., Lee, S. E., & Eom, B. (2017). Effects of age, working memory, and word order on passive-sentence comprehension: evidence from a verb-final language. *International Psychogeriatrics*, **29**(6), 939–948.

Thibault, M. P., Helloin, M. C., & Croteau, B. (2010). Exalang 5/8: Batterie informatisée pour l'examen du langage oral et écrit chez l'enfant de 5 à 8 ans, version 2. *Motus Éditions*.

Tomas, E., & Vissers, C. (2019). Behind the scenes of Developmental Language Disorder: time to call neuropsychology back on stage. *Frontiers in Human Neuroscience*, **12**, 517.

Tomasello, M. (2000). First steps toward a usage-based theory of language acquisition. *Cognitive Linguistics*, **11**(1/2), 61–82.

Tomasello, M. (2005). *Constructing a language*. Harvard University Press.

Tomasello, M. (2009). The usage-based theory of language acquisition. In *The Cambridge handbook of child language* (pp. 69–87). Cambridge University Press.

Tuller, L., Delage, H., Monjauze, C., Piller, A. G., & Barthez, M. A. (2011). Clitic pronoun production as a measure of atypical language development in French. *Lingua*, **121**, 423–441.

Tuller, L., Hamann, C., Chilla, S., Ferré, S., Morin, E., Prevost, P., & Zebib, R. (2018). Identifying language impairment in bilingual children in France and in Germany. *International Journal of Language and Communication Disorders*, **53**, 888–904.

Tuller, L., Henry, C., Sizaret, E., & Bathez, M. A. (2012). SLI at Adolescence: Avoiding Complexity. *Applied Psycholinguistics*, **33**, 161–184.

Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to language: the procedural deficit hypothesis. *Cortex*, **41**(3), 399–433.

Veer, I. M., Luyten, H., Mulder, H., van Tuijl, C., & Sleegers, P. J. (2017). Selective attention relates to the development of executive functions in 2, 5-to 3-year-olds: A longitudinal study. *Early Childhood Research Quarterly*, **41**, 84–94.

Vugs, B., Hendriks, M., Cuperus, J., & Verhoeven, L. (2014). Working memory performance and executive function behaviors in young children with SLI. *Research in Developmental Disabilities*, **35**(1), 62–74.

Vugs, B., Knoors, H., Cuperus, J., Hendriks, M., & Verhoeven, L. (2017). Executive function training in children with SLI: A pilot study. *Child Language Teaching and Therapy*, **33**(1), 47–66.

Wechsler, D. (2005). *WISC-IV : Echelle d'Intelligence de Wechsler pour Enfants – 4ème édition*. ECPA.

Weismer, S. E., Davidson, M. M., Gangopadhyay, I., Sindberg, H., Roebuck, H., & Kaushanskaya, M. (2017). The role of nonverbal working memory in morphosyntactic processing by children with specific language impairment and autism spectrum disorders. *Journal of Neurodevelopmental Disorders*, **9**(1), 28.

Weismer, S. E., Evans, J., & Hesketh, L. J. (1999). An examination of verbal working memory capacity in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, **42**(5), 1249–1260.

Weismer, S. E., Plante, E., Jones, M., & Tomblin, J. B. (2005). A functional magnetic resonance imaging investigation of verbal working memory in adolescents with specific language impairment. *Journal of Speech, Language, and Hearing Research*, **48**, 405–425.

Willis, C. S., & Gathercole, S. E. (2001). Phonological short-term memory contributions to sentence processing in young children. *Memory*, **9**, 349–364.

Zebib, R., Tuller, L., Hamann, C., Abed Ibrahim, L., & Prévost, P. (2019). Syntactic complexity and verbal working memory in bilingual children with and without Developmental Language Disorder. *First Language, Special Issue Article*, 1–24.

Zwitserlood, R., Wijnen, F., van Weerdenburg, M., & Verhoeven, L. (2015). 'MetaTaal': enhancing complex syntax in children with specific language impairment—a metalinguistic and multimodal approach. *International Journal of Language & Communication Disorders*, **50**(3), 273–297.