Language acquisition of early sequentially bilingual children is moderated by short-term memory for order in developmental language disorder: Findings from the HelSLI study

Lahti-Nuuttila, Pekka

2021-09

Lahti-Nuuttila, P., Laasonen, M., Smolander, S., Kunnari, S., Arkkila, E., & Service, E. (2021). Language acquisition of early sequentially bilingual children is moderated by short-term memory for order in developmental language disorder: Findings from the HelSLI study. International Journal of Language and Communication Disorders, 56(5), 907-926. https://doi.org/10.1111/1460-6984.12635

http://hdl.handle.net/10138/334325
https://doi.org/10.1111/1460-6984.12635

cc_by
publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.
Please cite the original version.
RESEARCH REPORT

Language acquisition of early sequentially bilingual children is moderated by short-term memory for order in developmental language disorder: Findings from the HelSLI study

Pekka Lahti-Nuuttila\textsuperscript{1,2} | Marja Laasonen\textsuperscript{1,2,3} | Sini Smolander\textsuperscript{1,4} | Sari Kunnari\textsuperscript{4} | Eva Arkkila\textsuperscript{1} | Elisabet Service\textsuperscript{2,5}

\textsuperscript{1} Department of Otorhinolaryngology and Phoniatrics, Head and Neck Surgery, Helsinki University Hospital and University of Helsinki, Helsinki, Finland
\textsuperscript{2} Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Helsinki, Finland
\textsuperscript{3} Logopedics, School of Humanities, Philosophical Faculty, University of Eastern Finland, Joensuu, Finland
\textsuperscript{4} Research Unit of Logopedics, University of Oulu, Oulu, Finland
\textsuperscript{5} Centre for Advanced Research in Experimental and Applied Linguistics (ARiEAL), Department of Linguistics and Languages, McMaster University, Hamilton, Canada

Abstract

Background: The role of domain-general short-term memory (STM) in language development remains controversial. A previous finding from the HelSLI study on children with developmental language disorder (DLD) suggested that not only verbal but also non-verbal STM for temporal order is related to language acquisition in monolingual children with DLD.

Aims: To investigate if a similar relationship could be replicated in a sample of sequentially bilingual children with DLD. In addition to the effect of age, the effect of cumulative second language (L2) exposure was studied.

Methods & Procedures: Sixty-one 4–6-year-old bilingual children with DLD and 63 typically developing (TD) bilingual children participated in a cross-sectional study conducted in their L2. Children completed novel game-like tests of visual and auditory non-verbal serial STM, as well as tests of cognitive functioning and language. Interactions of STM for order with age and exposure to L2 (Finnish) were explored as explanatory variables.

Outcomes & Results: First, the improvement of non-verbal serial STM with age was faster in sequentially bilingual TD children than in bilingual children with DLD. A similar effect was observed for L2 exposure. However, when both age and exposure were considered simultaneously, only age was related to the differential growth of non-verbal STM for order in the groups. Second, only in children with DLD was better non-verbal serial STM capacity related to an improvement in language scores with age and exposure.

Conclusions & Implications: The results suggest that, as previously found in Finnish monolingual children, domain-general serial STM processing is also compromised in bilingual children with DLD. Further, similar to the monolingual findings, better non-verbal serial STM was associated with greater language...
improvement with age and exposure, but only in children with DLD, in the age range studied here. Thus, in clinical settings, assessing non-verbal serial STM of bilingual children could improve the detection of DLD and understanding of its non-linguistic symptoms.

**KEYWORDS**
non-verbal serial short-term memory, language, vocabulary, developmental language disorder, specific language impairment, second language acquisition, sequentially bilingual, multilingual, memory for order

**What this paper adds**

What is already known on the subject
• Both phonological and non-verbal STM have been associated with DLD in monolingual and sequentially bilingual children. Monolingual children with DLD have also shown slower non-verbal serial STM development than TD children.

What this study adds to existing knowledge
• Sequentially bilingual TD children’s non-verbal serial STM improves more between ages 4 and 7 years than that of their peers with DLD, replicating a finding for monolingual children with DLD. Better non-verbal serial STM was especially associated with early receptive language development in sequentially bilingual children with DLD. L2 exposure showed largely comparable effects with age. These results support the hypothesis that a domain-general serial STM deficit is linked to DLD.

What are the potential or actual clinical implications of this work?
• Non-verbal assessment of STM for serial order in sequentially bilingual children with DLD could benefit the development of better tailored therapeutic interventions.

**INTRODUCTION**

The term developmental language disorder (DLD) has been proposed for ‘children who are likely to have language problems enduring into middle childhood and beyond, with a significant impact on everyday social interactions or educational progress’ (Bishop et al. 2017: 1070) but whose difficulties are not part of an identified biomedical condition. In contrast to the previously used label specific language impairment (SLI), DLD does not exclude deficits in non-verbal abilities and ‘children with low non-verbal ability who do not meet criteria for intellectual disability can be included as cases of DLD’ (Bishop et al. 2017: 1072). The present study introduces two non-linguistic tasks designed to tap short-term memory (STM) for order in time (also referred to as serial STM). These tasks are used to explore whether a domain-general mechanism for recording temporal serial order may be compromised in DLD. The earlier HelSLI study of monolingual 4–6-year-old children (Lahti-Nuuttila et al. 2021) suggested this was the case. This article presents data from a parallel sample of children acquiring a second language. Typically developing (TD) sequentially bilingual children and sequentially bilingual children with DLD were compared in a cross-sectional design to probe for differences in STM and language developmental patterns.

A number of impairments in non-linguistic cognitive processes have been associated with DLD, for example, in processing speed (Leonard et al. 2007), procedural learning (Ullman and Pierpont 2005) and sustained attention (Ebert and Kohnert 2011, Ebert et al. 2019, Finneran et al. 2009). Also, and more specifically, several recent studies and reviews have examined non-linguistic working memory (WM) as well as STM and how these are affected in DLD
DLD has mainly addressed two hypotheses identified by research on the relationship between STM and/or WM and DLD. Previous research on the relationship between STM and/or WM and DLD has mainly addressed two hypotheses identified by Vugs et al. (2013). The phonological storage deficit hypothesis of DLD (Archibald and Gathercole 2006a, 2006b, Baddeley et al. 1998, Gathercole and Baddeley 1990) suggests that mainly phonological WM is impaired in DLD. The alternative domain-general hypothesis of DLD asserts that general and non-verbal factors are involved in addition to phonological memory. A meta-analysis that examined the association between DLD and visuospatial WM found that children with DLD have deficits in both complex visuospatial WM tasks, in which information has to be actively processed as well as maintained, and simple visuospatial tasks (e.g., forwards and backwards digit span) and visuospatial tasks (e.g., forwards and backwards Corsi Blocks, which probes memory for ordered tapping of spatially distributed blocks or screen locations). In Corsi Blocks and similar tasks, temporal and spatial order are confounded as both memory for spatial and serial patterns affect performance. In the present study, the specific focus is on memory for temporal order. The question is whether a domain-general STM mechanism for order in time (as opposed to space) is related to DLD, and if it plays a role in atypical language development.

STM for verbal serial order, sometimes referred to as phonological STM, has been shown to predict typical vocabulary and grammar acquisition (e.g., Clark and Lum 2017, Hsu and Bishop 2011, Majerus and Boukebza 2013, Majerus et al. 2006b). A relation between phonological STM and DLD has also been reported in many investigations (Archibald 2017, Archibald and Gathercole 2006a, Baddeley 2003, Gathercole and Baddeley 1990, 1993, Montgomery et al. 2010, Verhagen and Leseman 2016). Furthermore, studies of monolingual TD children that examined STM for verbal items and their order separately found that these were independently linked to vocabulary acquisition (Attout et al. 2020, Leclercq and Majerus 2010, Majerus and Boukebza 2013, Majerus et al. 2006a, 2006b, Ordonez Magro et al. 2018). For example, in a recent study, a link between order STM and both receptive vocabulary and expressive vocabulary was found in 4–6-year-old TD children (Attout et al. 2020). In addition, in a study of 6–7-year-old TD children, better serial order reconstruction performance was related to faster novel word learning (Majerus and Boukebza 2013). In theories of STM, the nature of order coding mechanisms remains controversial. It has been suggested that order coding could be domain-general (Hurlstone et al. 2014), but recent research also points to the possibility of partly shared, partly domain-specific, or completely domain-specific mechanisms for verbal and non-verbal material (Hartley et al. 2016, Hurlstone 2019, Hurlstone and Hitch 2018).

Few studies have investigated STM for order in DLD. In a study including dyslexic children with or without DLD, Cowan et al. (2017) reported that children who had both DLD and dyslexia performed more poorly in serial order memory tasks than TD children. As the authors suggested, this might be explained by a deficit in general order memory. Because of the dearth of studies, the role of domain-general serial order STM in language acquisition remains unresolved. In a recent cross-sectional study in the HelSLI project (Lahti-Nuuttila et al. 2021), the associations between non-verbal serial STM and composite language measures of expressive language, receptive language and language reasoning were investigated in fifty-one 4–6-year-old monolingual Finnish children with DLD and 66 TD children. Non-verbal serial STM was found to improve more rapidly with age in the TD children than in the children with DLD. Furthermore, non-verbal serial STM (measured similarly as in the present study, that is, as a composite variable of non-verbal visual and auditory serial STM tasks) moderated the development of receptive language with age in the children with DLD but not in the TD group. Only in the children with DLD was better non-verbal serial STM related to better receptive language scores. Other studies that have found verbal or non-verbal STM impairment in children with DLD have mainly compared monolingual children with DLD with monolingual TD children. However, the relationship between memory for order of verbal material and language acquisition has also been found in bilingual children (e.g., Boerma et al. 2015, Engel de Abreu et al. 2014, Girbau and Schwartz 2008, Windsor et al. 2010). For example, Girbau and Schwartz (2008) compared children with DLD and TD children who had Spanish as a first language (L1) and English as a second language (L2) with a task thought to rely on memory for phoneme order, that is, a non-word repetition task with Spanish phonotactics. They found TD children to perform significantly better. Windsor et al. (2010) replicated this finding for L2 non-words.

In monolingual children, the general non-linguistic processing weaknesses (e.g., WM, sustained attention, processing speed) are linked to language acquisition and to DLD (Archibald 2017, Finneran et al. 2009, Ebert and Kohnert 2011, Leonard et al. 2007, Vugs et al. 2013). There has been much less research on bilingual children. However, a similar relationship has been suggested for bilingual children with language impairment (Kohnert et al. 2009, Kohnert 2010). A recent study replicated inferior non-verbal sustained attention and attentional control in bilingual as well as monolingual children with DLD compared with TD children (Ebert et al. 2019). Moreover, a
mediation analysis (Boerma et al. 2017) suggested an indirect role for sustained attention in the longitudinal development of vocabulary and morphology similarly in both mono- and bilingual groups of children with DLD despite different exposure rates to the tested language. To what extent other subclinical cognitive weaknesses, such as serial temporal order STM, interact with age and exposure in bilingual language difficulties is currently unknown.

The effects of age and exposure on language development cannot be separated in regular monolingual samples, but they are of interest for optimal targeting of interventions for specific groups with DLD. Studying L2 learners makes it possible to ask whether serial STM differently moderates the effects of age and cumulative L2 exposure on TD and DLD language performance. The current cross-sectional study investigates the specific hypothesis that domain-general serial STM moderates the language development of 4–6-year-old sequentially bilingual TD children and sequentially bilingual children with DLD. As in the earlier study of HeiSLI (Lahti-Nuuttila et al. 2021), domain-general STM moderation effects in relation to different aspects of language competence (expressive and receptive language as well as a broader domain of language reasoning tasks) were explored. The participants were 4–6-year-old early sequentially bilingual children who had acquired their L2, Finnish, between 0;1 and 5;10 years of age. If the associations of age, DLD and non-verbal serial STM with language turned out to be similar in bilingual children, as found in the monolingual children of the earlier HeiSLI study, this would suggest that assessment of non-verbal serial STM could be informative for identifying DLD in young bilingual children.

The assessment of non-verbal serial STM in the current study was designed to make minimal demands on proficiency in L2 as the task instructions were straightforward and the child could respond non-verbally. When optimized for sensitivity and specificity in young children, such a task could also be helpful for testing children with limited L2 exposure when testing in their L1 is not feasible. Furthermore, the functioning of non-verbal serial STM may relate to specific limitations of information processing, particularly of building memory representations for structure in time. Such temporal structure processing is central to learning the phonological structures of words and combination of words to phrases and sentences. Better understanding for these processes could, thus, inform interventions for DLD.

Based on the conception that memory for order is necessary for language acquisition, it was hypothesized that bilingual children with DLD have poorer and more slowly improving non-verbal serial STM capacity than bilingual TD children. Another hypothesis was that the development of language competence with age and L2 exposure is moderated by the development of non-verbal serial STM capacity. If serial STM capacity growth is different in children with DLD and TD children, cross-sectionally studied language development could also be differently moderated by STM in TD children compared with children with DLD. Therefore, it is hypothesized that significant interactions between participant group (TD versus DLD), age or L2 exposure, and serial STM in predicting composite language variables will be revealed.

**METHODS**

**Participants**

The group of sequentially bilingual children with DLD consisted of 61 children (46 boys) and the sequentially bilingual TD group of 63 children (47 boys). All children were between the ages of 4;0 and 7;3 (mean = 5;7, SD = 0;10). All had only one language other than Finnish as their L1, but there were 33 different L1s (see table S1 in the additional supporting information). Finnish was the only L2 with at least 7 months of exposure (mean = 3;0, SD = 1;3), and all but four children with DLD had more than 1 year of exposure. The mean age of onset for L2 was 2;7, SD = 1;1. None of the children had any gross neurological difficulties (e.g., diagnoses of autism spectrum disorder (ASD), epilepsy or chromosomal abnormalities), hearing impairment, intellectual disability or oral anomalies. Parental consent was obtained for each child participating in the study. Ethical approval for the study had been granted by the ethical board of the Hospital District of Helsinki and Uusimaa.

The children with DLD had been referred to the Auditory-Phoniatric Ward for Children, Department of Phoniatrics, Helsinki University Hospital for suspected DLD. They were examined during their visits to the ward and were diagnosed with ICD-10 (WHO 2010) as having a language disorder. Diagnoses of other developmental disorders (e.g., hearing impairment, intellectual disability, ASD, oral anomalies, or a diagnosed neurological impairment or disability) were used as exclusion criteria. Non-verbal intelligence was also part of the exclusion criteria, and a performance intelligence quotient (PIQ) of at least 70 was a requisite for inclusion for children with DLD. A total of 21 of the recruited children with DLD had a PIQ between 70 and 84 based on the Wechsler Preschool and Primary Scale of Intelligence—Third Edition (WPPSI-III) (Wechsler 2009). In the final sample, the PIQ of TD children (mean = 101.0, SD = 11.5) was statistically significantly higher than the PIQ of children with DLD (mean = 92.5, SD = 14.8) (p < 0.001, d = 0.64), which is in line with the results of the meta-analysis of Gallinat and Spaulding.
The present sample is representative of the DLD children generally assessed in the ward, and their inclusion follows statement 8 of the recent Criteria and Terminology Applied to Language Impairments: Synthesising the Evidence (CATALISE) consensus report (Bishop et al. 2017), which acknowledges that children with DLD can have low levels of non-verbal ability.

The bilingual TD children were voluntary participants from kindergartens in the metropolitan area of Helsinki. They were required not to have any diagnosed or suspected language difficulties except possible minor articulation impediments. TD children were required to have a PIQ of at least 85. Before CATALISE (Bishop et al. 2017), the initial plan had been to split the DLD group into two subgroups at PIQ = 85. After the CATALISE consensus process, the terminology and criteria were revised. It was also found that splitting the DLD group would have resulted in unacceptably small samples for the planned analyses. Consequently, the children with DLD were included as one group, and non-verbal reasoning was statistically controlled. It can be noted that in the initial screening of data the relationships between non-verbal subtests (see the section ‘Language and cognitive tests’) were very similar throughout the whole PIQ range.

Estimates of exposure to L2 were obtained from the Finnish version of the Alberta Language Environment Questionnaire (ALEQ) (Paradis 2011, Smolander et al. 2021). First, the number of months between the age at which the child began to have regular kindergarten exposure to Finnish and the age at which they participated in the present study was calculated. Based on the questions of ALEQ addressing the proportion of L1 and L2 languages in the child’s life, a cumulative L2 exposure score was then calculated as a product of L2 proportion and L2 exposure (Smolander et al. 2021). According to the information gained with ALEQ, the most important source of L2 exposure was the Finnish kindergarten, but also interaction with family members and peers, hobbies and other activities were taken into account. For an even more detailed description of the participants and more precise criteria related to exposure and other inclusion/exclusion criteria, see Laasonen et al. (2018); for a more comprehensive report about the estimate of L2 exposure, see Smolander et al. (2021).

Descriptive statistics for both groups are shown in table 1. The clinical context and the young age of the participants resulted in missing values in some language and cognitive tests. A total of 11 children had one missing value and five children had two. Missing value frequencies are reported in table 1. The groups’ ages did not significantly differ. Neither did L2 exposure differ significantly between TD children and children with DLD. Children in the TD group had significantly higher scores than the DLD group in all language tests with large effect sizes in 15 of 17 comparisons. They also had higher absolute scores in the non-verbal tests, most differences being statistically significant, although the effect sizes were smaller than in the verbal tests. To control potential confounds, age, L2 exposure and non-verbal test differences were adjusted using a propensity score method. Propensity-score adjusted standardized mean differences are presented in table 1.

Language and cognitive tests

The children had 33 different first languages. Since the focus of the present study was L2 acquisition, children were assessed in Finnish, their L2. Finnish is a morphologically complex agglutinating language in which most tokens of nouns, verbs and adjectives are inflected forms, consisting of two or multiple morphemes. Finnish is not closely related to other major languages except Estonian (and more distantly Hungarian). The choice of testing measures was limited to those that have been standardized for use in Finland. Picture Naming, Receptive Vocabulary, Information, Vocabulary, Word Reasoning, Block Design and Matrix Reasoning were selected from the WPPSI-III (Wechsler 2009). The Comprehension of Instructions, Imitating Hand Positions, Theory of Mind (Contextual Task) and Design Copying subtests were selected from the Nepsy-II (Korkman et al. 2008). In addition, the Comprehension and Expressive Scales subtests of the Reynell Developmental Language Scales III (Edwards et al. 1997) were administered. Children were also assessed using the Expressive (Martin and Brownell 2011) and Receptive One-Word Picture Vocabulary Tests (Martin and Brownell 2010) as well as the Boston Naming Test (Kaplan et al. 1983). The raw scores of these variables, sample-centred transformations of the raw scores and sample-standardized z-transformations of raw scores were used when appropriate in the particular analyses (for a description of the roles of the variables, see the section ‘Statistical analyses’).

Serial STM tasks

Two serial STM tasks were developed to test immediate memory for temporal order in non-verbal sequences. The STM tasks were presented to the child as tablet computer games. Pictures of four barns were shown on the screen. Two opposing upper barns were described as belonging to Matt and two lower barns to Mary. In both auditory and visual STM tasks, lengthening pairs of stimulus sequences were presented for comparison of order. In both modalities, the participants had to bind the stimuli, presented one at a time, to a temporal sequence in their WM.
TABLE 1  Descriptive statistics of fundamental variables by group and the results for mean comparisons

| Variable                        | Group   | TD (n = 63) | N of missing | Range   | DLD (n = 61) | N of missing | Range   | p     | d<sup>b</sup> | PS adjusted d<sup>b</sup> |
|---------------------------------|---------|-------------|--------------|---------|--------------|--------------|---------|-------|----------|--------------------------|
| Age (months)                    |         | 67.0(10.6)  | –            | 49–87   | 66.7(9.3)    | –            | 48–83   | 0.850 | 0.03     | 0.00                     |
| Cumulative L2 exposure (months) |         | 18.5 (8.3)  | –            | 7–38    | 17.2 (8.1)   | –            | 3–40    | 0.426 | 0.14     | 0.00                     |
| Non-verbal Reasoning<sup>c,d</sup> |        | 0.2 (0.9)   | –            | (−2.0)–2.2 | (−0.2) (0.9) | –            | (−1.7)–2.4 | 0.033 | 0.38     | 0.00                     |
| Matrix Reas.<sup>c</sup>        |         | 16.7 (4.3)  | –            | 5–25    | 14.5 (4.6)   | –            | 4–24    | 0.006 | 0.49     | 0.00                     |
| Block Design<sup>c</sup>        |         | 27.3 (4.7)  | –            | 20–38   | 26.4 (3.9)   | –            | 20–40   | 0.255 | 0.20     | 0.00                     |
| Imit. Hand Pos.<sup>c</sup>     |         | 15.2 (4.4)  | 1            | 5–23    | 10.7 (5.0)   | 2            | 2–20    | <0.001 | 0.93     | 0.00                     |
| Theory of Mind<sup>e</sup>      |         | 4.9 (1.7)   | –            | 0–8     | 4.1 (1.5)    | –            | 1–8     | 0.009 | 0.47     | 0.00                     |
| Design Copying<sup>e</sup>      |         | 7.2 (2.5)   | 1            | 3–14    | 5.8 (1.8)    | –            | 2–11    | <0.001 | 0.63     | 0.00                     |
| Vocab.<sup>c</sup>              |         | 14.4 (6.7)  | –            | 3–33    | 6.4 (4.0)    | –            | 0–18    | <0.001 | 1.45     | 1.06                     |
| Inform.<sup>c</sup>             |         | 21.8 (4.3)  | –            | 4–30    | 13.8 (6.4)   | –            | 0–23    | <0.001 | 1.47     | 0.99                     |
| Word Reas.<sup>c</sup>          |         | 12.8 (7.1)  | 1            | 1–23    | 3.5 (4.5)    | –            | 0–16    | <0.001 | 1.54     | 1.14                     |
| Compr. Instr.<sup>e</sup>       |         | 17.3 (4.7)  | –            | 5–28    | 11.6 (4.6)   | –            | 0–23    | <0.001 | 1.23     | 0.85                     |
| BNT                             |         | 16.0 (6.4)  | –            | 2–30    | 8.7 (5.9)    | 6            | 0–29    | <0.001 | 1.15     | 0.83                     |
| RDLS Expr.                      |         | 31.0 (12.2) | –            | 5–48    | 15.8 (8.7)   | 7            | 0–37    | <0.001 | 1.35     | 1.01                     |
| EOWPVT                          |         | 48.0 (15.6) | –            | 15–83   | 29.6 (15.3)  | 1            | 0–74    | <0.001 | 1.18     | 0.86                     |
| Pict. Naming<sup>c</sup>        |         | 15.5 (4.0)  | –            | 6–22    | 9.4 (5.6)    | –            | 0–20    | <0.001 | 1.25     | 0.93                     |
| Receptive Voc.<sup>c</sup>      |         | 23.9 (5.4)  | –            | 11–32   | 19.4 (7.6)   | –            | 1–33    | <0.001 | 0.68     | 0.36                     |
| RDLS Compr.                     |         | 50.7 (8.0)  | 16–61        | 38.3 (14.8) | – | 0–57           | <0.001 | 1.05 | 0.68     |                         |
| ROWPVT                          |         | 77.8 (30.0) | 28–158       | 43.8 (18.1) | 2 | 8–88           | <0.001 | 1.34 | 0.80     |                         |

Note: TD, typically developing children; DLD, children with developmental language disorder; d, Cohen's d, effect size; PS adjusted d, propensity score adjusted Cohen's d, effect size; Matrix Reas., Matrix Reasoning; Imit. Hand Pos., Imitating Hand Positions; Vocab., Vocabulary; Inform., Information; Word Reas., Word Reasoning; Compr. Instr., Comprehension of Instructions; BNT, Boston Naming Test; RDLS, Reynell Developmental Language Scales III; Expr., Expressive Scale; Compr., Comprehension Scale; EOWPVT, Expressive One Word Picture Vocabulary Test; Pict. Naming, picture naming; Receptive Voc., receptive vocabulary; and ROWPVT, Receptive One Word Picture Vocabulary Test.

<sup>a</sup> In the case of missing values, p- and d-values were pooled from the independent samples t-tests in 20 multiple imputations.

<sup>b</sup> Propensity scores were based on age, cumulative L2 exposure and the raw scores of Matrix Reasoning, Block Design, Imitating Hand Positions, Theory of Mind and Design Copying.

<sup>c</sup> Wechsler Preschool and Primary Scale of Intelligence, Third edition (Wechsler 2009).

<sup>d</sup> Non-verbal reasoning score is the mean of sample standardized z-scores of Matrix Reasoning and Block Design raw scores.

<sup>e</sup> Nepsy-II (Korkman et al. 2008).
TABLE 2  Serial STM tasks and language composite scores: Descriptive statistics by group and the results for mean comparisons

| Variable                  | Group               | TD (n = 63) | DLD (n = 61) | p     | d<sup>a</sup> |
|---------------------------|---------------------|-------------|--------------|-------|--------------|
|                           | Mean (SD)           | Mean (SD)   | Range        |       |              |
| **Serial short-term memory** |                     |             |              |       |              |
| Visual serial STM<sup>b</sup> | 6.9 (6.2)           | 4.6 (3.3)   | 0–24         | 0.010 | 0.46         |
| Auditory serial STM<sup>b</sup> | 8.0 (7.2)           | 4.1 (3.5)   | 1–24         | <0.001| 0.69         |
| Serial STM composite      | 0.3 (1.0)           | −0.3 (0.4)  | (−0.8)–3.6   | <0.001| 0.69         |
| **Language**              |                     |             |              |       |              |
| General language composite | 0.5 (0.7)           | −0.5 (0.7)  | (−1.1)–1.8   | <0.001| 1.46         |
| Expressive language composite | 0.5 (0.8)      | −0.5 (0.8)  | (−1.2)–1.8   | <0.001| 1.38         |
| Receptive language composite | 0.4 (0.7)           | −0.5 (0.8)  | (−1.2)–1.7   | <0.001| 1.17         |
| Language reasoning composite | 0.6 (0.8)       | −0.6 (0.7)  | (−1.1)–2.2   | <0.001| 1.63         |

Notes: TD, typically developing children; DLD, children with developmental language disorder; d, Cohen’s d, effect size; and STM, short-term memory.

<sup>a</sup>P- and d-values are pooled from the independent samples t-tests in 20 multiple imputations.

<sup>b</sup>One TD and one DLD child had visual serial STM task missing. One other TD child had a missing value for the auditory serial STM task. Three high score values in both STM tasks were winsorized in the TD group.

In the visual task, a first sequence of fantasy animals travelled one by one from Matt’s left barn to his right barn. After a short pause, a second sequence of animals moved from Mary’s left barn to her right barn. Each sequence consisted of tokens of two different animals sampled from the pool of five possible animals. Matt’s and Mary’s paired sequences always had the same two animals. After each pair of sequences, the child had to touch a green circle with a tick mark on the screen if Mary’s animals had moved in the same order as Matt’s, and a red circle with a cross if they had appeared in a different order.

In the auditory task, tokens of two different back-to-front animal calls sampled from the pool of five possible calls were used on each trial. In this task, Matt’s and Mary’s barns were seen as in the visual task, but now it was evening dusk. No animals were visible, but their calls could be heard. Matt’s right-side barn was lit during each call in the first sequence of sounds as invisible animals moved in and said good night. Mary’s right-side barn was lit during each call in the second sequence. Again, the child was asked to check whether the sequences were identical.

In half the comparisons at each sequence length, Matt’s and Mary’s sequences were the same, and in the other half, they were different. First, five practice comparisons were presented to make sure that the child had understood the task. In the actual task, six comparisons per sequence length were presented. The initial sequence length was two, and it increased only if the child responded correctly on at least four out of six trials of the current length. If the child responded correctly on the first four trials, the last two trials of that sequence length were not presented but were credited. The children's score in each task was the number of actual correct answers and these credits. The maximum sequence length was seven, so the theoretical maximum score was 36. Half the children were presented with the auditory task first, whereas the other half was first presented with the visual task.

For practical reasons, only a limited number of trials could be included in the tasks. For a more reliable STM measure, the visual and auditory scores were, therefore, standardized, and a composite STM score was calculated as the average of the standard scores. The combination of visual and auditory tasks also served to control for modality-specific strategies. The descriptive statistics for the STM and the language composite variables used in the main analyses are presented in table 2.

**STATISTICAL ANALYSIS**

The main goal of the present study was to examine the relationship between non-verbal STM for order and language development in bilingual TD children and bilingual children with DLD as a function of age and exposure. From the 11 observed language variables, composites were formed for receptive language, expressive language and language reasoning (cf., Lahti-Nuuttila et al. 2021), as well as a second-order composite variable for general language. A receptive language composite was formed as a mean of sample standardized values of the Reynell III Comprehension Scale, the Receptive One-Word Picture Vocabulary Test and Receptive Vocabulary of WPPSI-III. The expressive language composite included sample standardized values of the Reynell III Expressive Scale, the Expressive One-Word Picture Vocabulary Test, the Boston Naming Test and Picture Naming from WPPSI-III. The remainder of the language tests, that is, Information, Vocabulary...
and Word Reasoning from WPPSI-III and the Comprehension of Instructions from the Nepsy-II, formed the language reasoning composite. A general language composite was formed as an average of receptive and expressive language and language reasoning composites.

The initial screening of the data revealed that there were slightly fewer than expected raw scores at the high end of the distribution on some cognitive measures among the children with DLD who were over 5.6 years of age: 5.5-year-olds had equivalent raw scores to many 4-year-old children. The local regression (loess) curves confirmed that some older children with DLD performed relatively worse than the younger children with DLD. To correct for this possible slight bias resulting from an unequal age distribution of cognitive skills in the DLD group, and to adjust for the group difference in non-verbal reasoning, propensity scores were used (Rosenbaum and Rubin 1983, Schafer and Kang 2008).

A propensity score is a balance score (Austin 2011) that can control for possible confounding that results from unintended group differences. The good thing about the propensity score method is that it can control for many possible confounders at the same time. One way to create a propensity score for a measure is to employ logistic regression to predict group membership with a set of explanatory variables that may need to be controlled. A propensity score estimate is based on the predicted probability of group membership found in this analysis. The propensity score can be used to create propensity score classes (Schafer and Kang 2008). In the current study, a propensity score analysis was conducted with binary logistic regression, using the group as the outcome variable, and age, cumulative L2 exposure, as well as the raw scores of the non-verbal measures Matrix Reasoning, Block Design, Imitating Hand Positions, Theory of Mind and Design Copying as predictor variables. The propensity scores were estimated as the predicted group membership probabilities. Balance checking between groups was conducted with these propensity scores. For the regression analyses of non-verbal serial STM and the language composite variables, a procedure proposed by Schafer and Kang (2008) was used. Subjects were classified into five propensity score classes, and four dummy variables that distinguished these classes were used as covariates in the main analyses. The dummy variables were constructed so that all observations that were classified as belonging to the first propensity score class received a value of 1 and all other observations received a value of 0 for the first dummy variable. The three other dummy variables were coded similarly. These variables were used as covariates in the regression models of interest.1

For the missing values in the data (table 1), the multiple imputation (MI) procedure of SPSS 25 with 20 imputed data sets was used. The results are reported pooled and based on small-sample degrees of freedom (Reiter 2007, van Ginkel and Kroonenberg 2014). The MI was performed for the raw scores of the language, cognitive and STM variables before centring or standardizing with gender, group status, age, L2 exposure and cumulative L2 exposure also in the MI model. In the TD group, three high outlier values in both the visual and auditory STM tasks were detected. These raw values were winsorized before calculating the non-verbal serial STM composite so that they would not disproportionately influence the results.

The main interest was moderator effects, in other words, interactions. To make the interpretation of the effect estimates more comprehensible, the predictor variables were mean-centred for estimating unstandardized effects, following common practice in moderation analyses (Hayes and Rockwood 2017). The standardized effects (β-coefficients) were estimated with sample standardized (z-transformed) variables using the GLM procedure of SPSS 25.0.0.2. In the analyses where the STM composite was the dependent variable, the statistically significant interaction of age and group was further cross-checked using the PROCESS macro (Hayes 2018), and tests of conditional effects (effects of age within the groups) were estimated with the macro-procedures separately for each multiply imputed sample. Again, the results from all 20 samples were pooled using small-sample degrees of freedom (Reiter 2007, van Ginkel and Kroonenberg 2014). Two-tailed statistical significance tests were used, and the significance level was originally set as α = 0.05.

RESULTS

Balance check of propensity scores

Standardized mean differences between the groups using the propensity score as a covariate in the analyses are presented in table 1. These show that for age, for L2 cumulative exposure and for the non-verbal reasoning and its subtests (Matrix Reasoning and Block Design), the groups were balanced, all d’s being near 0 and the group differences nonsignificant. In the classification of propensity scores (Schafer and Kang 2008), the numbers of TD children in the five bins were 22, 20, 9, 10 and 1, while the numbers of children with DLD were 2, 5, 16, 15 and 24, respectively.

Relationship of age and cumulative L2 exposure with serial STM

Correlations between age, L2 cumulative exposure, non-verbal reasoning composite, serial STM composite and
TABLE 3  Predicting non-verbal serial short-term memory: Results of the multiple regression analyses

| Model 1 |   |   |   |   |   |
|---------|---|---|---|---|---|
|         | B | SE | 95% CI | p  | β  | R² |
| Age     | 0.03 | 0.006 | [0.02, 0.04] | <0.001 | 0.37 |
| DLD     | -0.49 | 0.154 | [-0.80, -0.19] | 0.002 | -0.29 |
| Age × DLD | -0.05 | 0.012 | [-0.08, -0.03] | <0.001 | -0.33 |

| Model 2 |   |   |   |   |   |
|---------|---|---|---|---|---|
|         | B | SE | 95% CI | p  | β  | R² |
| Age     | 0.03 | 0.008 | [0.02, 0.05] | <0.001 | 0.31 |
| DLD     | -0.47 | 0.170 | [-0.81, -0.13] | 0.007 | -0.28 |
| Age × DLD | -0.04 | 0.016 | [-0.08, -0.01] | 0.007 | 0.22 |

| Model 3 |   |   |   |   |   |
|---------|---|---|---|---|---|
|         | B | SE | 95% CI | p  | β  | R² |
| Age     | 0.03 | 0.007 | [0.01, 0.04] | <0.001 | 0.32 |
| Cumulative L2 exposure | 0.01 | 0.008 | [0.00, 0.03] | 0.099 | 0.13 |
| DLD     | -0.46 | 0.161 | [-0.78, -0.14] | 0.005 | -0.28 |
| Age × DLD | -0.05 | 0.014 | [-0.08, -0.03] | <0.001 | -0.32 |
| Cumulative L2 exposure × DLD | -0.01 | 0.017 | [-0.04, 0.02] | 0.617 | -0.04 |
| Age × Cumulative L2 exposure | 0.00 | 0.001 | [0.00, 0.00] | 0.129 | 0.12 |
| Age × Cumulative L2 exposure × DLD | -0.00 | 0.002 | [0.00, 0.00] | 0.302 | -0.08 |

Note: DLD, dummy variable, which before centring had 0 = typically developing children and 1 = children with developmental language disorder and after centring −0.49 and 0.51, respectively; STM, short-term memory. Also, four propensity score class dummy variables were included in the model to adjust TD and DLD group differences, but reporting their coefficients and that of the intercept is not relevant.

*p-values were calculated using small-sample degrees of freedom for multiple imputations (Reiter 2007; Van Ginkel and Kroonenberg 2014).

FIGURE 1  Serial STM composite score by age × group (left) and by cumulative L2 exposure × group (right). The thin dashed lines show regression of the x-variable only, while the thick solid lines present estimations of models 1 and 2 from table 3 where possibly confounding group differences were controlled via propensity score groups. Thick solid lines are for classified propensity score = 3

language composites are presented in table S2 in the additional supporting information. The model with age, group, age × group and propensity score class dummies predicting serial STM was statistically significant ($F_{7,114} = 13.4, p < 0.001$) and explained 45% of the variance in the non-verbal serial STM composite. The regression coefficients for age, group and age × group are presented in table 3. The four propensity score class dummy variables as a combination also had a significant effect ($p = 0.009$). The age × group interaction (figure 1) was statistically significant, suggesting that the cross-sectionally obtained effect of age on serial STM was different in children with DLD than TD children, only TD children showing serial STM improvement with age. A follow-up analysis of the interaction in model 1 showed that conditionally for group ($b_{\text{cond. Age}} = b_{\text{Age}} + b_{\text{Age × group × centred group value}}$) the age effect in the TD group ($b = 0.058, p < 0.001$) was significant while in the DLD group it was not ($b = 0.003, p = 0.710$). When cumulative L2 exposure was used as a predictor in the model instead of age, the model was also
significant (\(F_{7,114} = 7.7, p < 0.001\)) but not as good as model 1 (table 3).

Finally, a model with both age and cumulative L2 exposure and their interactions with group (table 3, model 3) had a slightly higher coefficient of determination (\(F_{11,110} = 10.0, p < 0.001\)). In this model, the only significant interaction was age \(\times\) group, with a comparable \(\beta\)-coefficient to model 1. Therefore, it seems that the age \(\times\) group interaction was similar but stronger than the cumulative L2 exposure \(\times\) group interaction. The STM composite score of TD children increased with age and exposure, while that of children with DLD did not show significant improvement with either predictor.

**Age, DLD and serial STM as predictors of language composites**

The results from the regression analyses of the models where each language composite was regressed on age, group, serial STM and their interactions are summarized in table 4.

When predicting the general language composite with the model including propensity score class, age, group status, age \(\times\) group, serial STM, age \(\times\) serial STM, group \(\times\) serial STM and age \(\times\) group \(\times\) serial STM, the model was statistically significant (\(F_{11,110} = 24.1, p < 0.001\)). The main effects of age and group were statistically significant, as were the two-way interactions of age \(\times\) group and age \(\times\) serial STM. Most importantly, the three-way interaction of age \(\times\) group \(\times\) serial STM on general language was significant. Thus, the role of STM appears different for children with DLD than for TD children. This pattern was also seen in the separate analyses of the three first-order language composites that are presented below and pictured in figure 2, showing estimated developmental projections in the 20th, 50th and 80th percentiles of STM performance. Children with DLD who had higher STM composite scores were found to have a steeper language growth than children with DLD and lower STM composite scores. In TD children, higher or lower STM performance does not seem to associate with language development.

The model for the expressive language composite was statistically significant (\(F_{11,110} = 17.5, p < 0.001\)) with significant main effects of age and group, significant two-way interactions of age \(\times\) group and age \(\times\) serial STM, and a significant three-way interaction of age \(\times\) group \(\times\) serial STM (table 4).

Similarly, the model for the receptive language composite was statistically significant (\(F_{11,110} = 18.2, p < 0.001\)), including significant main effects of age and group, significant two-way interactions of age \(\times\) group and age \(\times\) serial STM, two-way interaction of age \(\times\) serial STM and a significant three-way interaction of age \(\times\) group \(\times\) serial STM (table 4).

Lastly, the model predicting the language reasoning composite was statistically significant (\(F_{11,110} = 28.0, p < 0.001\)), including the significant main effects of age and group, and a marginally significant three-way interaction of age \(\times\) group \(\times\) serial STM (\(p = 0.051\)). However, the age \(\times\) group interaction was not significant (table 4).

The models for the three language composites are illustrated in figure 2. For probing the interactions, three percentile values (20th, 50th and 80th) of the serial STM composite were chosen for the figure. The lines in the figure are for the middle (third) propensity score class. These suggest that those children with DLD who have better STM show greater language improvement with age than children with DLD who have poorer STM capacities, whereas serial STM capacity does not seem to predict language development for TD children in this age/skill range.

**Cumulative L2 exposure, DLD and serial STM as predictors of language composites**

Exposure to language is perfectly confounded with age in monolingual children. However, bilingual children’s L2 exposure is somewhat separate from age. This allows the examination of exposure as a predictor variable. The results from the regression analyses of the models where each language composite was regressed on cumulative L2 exposure, group, serial STM and their interactions are shown in table 5. Here, all the models significantly predicted the language composites (general language composite: \(F_{11,110} = 15.3, p < 0.001\); expressive language composite: \(F_{11,110} = 14.3, p < 0.001\); receptive language composite: \(F_{11,110} = 10.7, p < 0.001\); language reasoning composite: \(F_{11,110} = 16.2, p < 0.001\)). The main effects of cumulative L2 exposure and group were significant in every model. In all of these models, except in the model for receptive language, the three-way interaction L2 exposure \(\times\) group \(\times\) serial STM effect was also significant.

**Age, cumulative L2 exposure, DLD, serial STM and language composites**

Finally, it was attempted to deconfound age and cumulative exposure by analysing two sets of models where both age and cumulative L2 exposure were regressors, but interactions for only one of these variables were included. For balanced comparison, standardized variables were analysed. Results are shown in table 6. All the models that included interactions with age were significant (general language composite: \(F_{12,109} = 28.0, p < 0.001\); expressive
TABLE 4  Predicting language composites: Results of the multiple regression analyses with centred age (months), group status, non-verbal serial short-term memory and their interactions as predictors

|                                      | B     | SE    | 95% CI     | \(p^*\) | \(\beta\) | \(R^2\) |
|--------------------------------------|-------|-------|------------|---------|---------|---------|
| **General language composite**       |       |       |            |         |         | 0.72    |
| Age                                  | 0.06  | 0.006 | [0.05, 0.07]| <0.001  | 0.68    |         |
| DLD                                  | -0.97 | 0.135 | [-1.24, -0.70]| <0.001  | -0.56   |         |
| Age × DLD                            | 0.03  | 0.013 | [0.00, 0.05]| 0.041   | 0.16    |         |
| Non-verbal serial STM                | -0.06 | 0.099 | [-0.25, 0.14]| 0.569   | -0.05   |         |
| Age × Non-verbal serial STM          | 0.03  | 0.012 | [0.01, 0.05]| 0.008   | 0.30    |         |
| DLD × Non-verbal serial STM          | -0.19 | 0.198 | [-0.58, 0.20]| 0.335   | -0.09   |         |
| Age × DLD × Non-verbal serial STM    | 0.06  | 0.024 | [0.01, 0.11]| 0.017   | 0.28    |         |
| **Expressive language composite**    |       |       |            |         |         | 0.65    |
| Age                                  | 0.06  | 0.008 | [0.04, 0.07]| <0.001  | 0.63    |         |
| DLD                                  | -0.96 | 0.159 | [-1.28, -0.65]| <0.001  | -0.53   |         |
| Age × DLD                            | 0.03  | 0.016 | [0.00, 0.06]| 0.038   | 0.18    |         |
| Non-verbal serial STM                | -0.00 | 0.116 | [-0.23, 0.23]| 0.974   | -0.00   |         |
| Age × Non-verbal serial STM          | 0.04  | 0.014 | [0.01, 0.06]| 0.010   | 0.32    |         |
| DLD × Non-verbal serial STM          | -0.09 | 0.233 | [-0.55, 0.38]| 0.713   | -0.04   |         |
| Age × DLD × Non-verbal serial STM    | 0.06  | 0.028 | [0.01, 0.12]| 0.031   | 0.28    |         |
| **Receptive language composite**     |       |       |            |         |         | 0.65    |
| Age                                  | 0.06  | 0.007 | [0.05, 0.08]| <0.001  | 0.69    |         |
| DLD                                  | -0.82 | 0.152 | [-1.12, -0.52]| <0.001  | -0.46   |         |
| Age × DLD                            | 0.03  | 0.015 | [0.01, 0.06]| 0.021   | 0.20    |         |
| Non-verbal serial STM                | -0.07 | 0.111 | [-0.29, 0.15]| 0.545   | -0.06   |         |
| Age × Non-verbal serial STM          | 0.03  | 0.013 | [0.00, 0.05]| 0.040   | 0.25    |         |
| DLD × Non-verbal serial STM          | -0.30 | 0.222 | [-0.74, 0.14]| 0.184   | -0.14   |         |
| Age × DLD × Non-verbal serial STM    | 0.07  | 0.027 | [0.01, 0.12]| 0.015   | 0.31    |         |
| **Language reasoning composite**     |       |       |            |         |         | 0.74    |
| Age                                  | 0.06  | 0.006 | [0.05, 0.07]| <0.001  | 0.63    |         |
| DLD                                  | -1.12 | 0.134 | [-1.39, -0.86]| <0.001  | -0.62   |         |
| Age × DLD                            | 0.01  | 0.013 | [-0.01, 0.04]| 0.260   | 0.08    |         |
| Non-verbal serial STM                | -0.10 | 0.098 | [-0.29, 0.10]| 0.318   | -0.09   |         |
| Age × Non-verbal serial STM          | 0.03  | 0.011 | [0.01, 0.05]| 0.007   | 0.29    |         |
| DLD × Non-verbal serial STM          | -0.19 | 0.196 | [-0.58, 0.20]| 0.329   | -0.09   |         |
| Age × DLD × Non-verbal serial STM    | 0.05  | 0.024 | [0.00, 0.09]| 0.051   | 0.21    |         |

Note: DLD, dummy variable, which before centring had 0 = typically developing children and 1 = children with developmental language disorder and after centring -0.49 and 0.51, respectively; STM, short-term memory. Also, four propensity score class dummy variables were included in the model to adjust TD and DLD group differences but reporting their coefficients is irrelevant.

*P*-values were calculated using small-sample degrees of freedom for multiple imputations (Reiter 2007; Van Ginkel and Kroonenberg 2014).

language composite: \(F_{12,109} = 21.6, p < 0.001\); receptive language composite: \(F_{12,109} = 19.8, p < 0.001\); language reasoning composite: \(F_{12,109} = 29.8, p < 0.001\). The three-way interaction effect of age \(\times\) group \(\times\) serial STM was significant only when predicting general language and receptive language, although it showed non-significant trends also for the other two composites (\(p = 0.054\) for expressive language and \(p = 0.085\) for language reasoning).

The models with cumulative L2 exposure interactions were likewise significant (general language composite: \(F_{12,109} = 28.9, p < 0.001\); expressive language composite: \(F_{12,109} = 21.9, p < 0.001\); receptive language composite: \(F_{12,109} = 19.9, p < 0.001\); language reasoning composite: \(F_{12,109} = 30.2, p < 0.001\)). The three-way interaction effect of cumulative L2 exposure \(\times\) group \(\times\) serial STM was significant in all but in the model for receptive language (\(p = 0.097\)).

Taken together, these results show that for children with DLD, the language boosting effect of better non-verbal STM was reliably detectable on receptive language as a function of age. For expressive language and
language reasoning, the boosting effect was detectable for language improvement as a function of cumulative L2 exposure. However, if non-significant trends are considered, the results for age and cumulative L2 exposure were similar. For TD children, who had better STM than children with DLD, the boosting effect was not found, but their language scores improved simply as a function of age and cumulative L2 exposure.

**Discussion**

Studies on the role of STM and WM in DLD have mainly concentrated on phonological STM and verbal WM as possible causes of DLD (e.g., Archibald and Gathercole 2006a, Gathercole and Baddeley 1990; for a review, see Archibald 2017). Investigations of non-verbal memory have reported findings of deficient visuo-spatial executive WM
TABLE 5 Predicting language composites: Results of the multiple regression analyses with centred L2 cumulative exposure, group status, non-verbal serial short-term memory and their interactions as predictors

|                                | B    | SE  | 95% CI        | p   | β   | R²  |
|--------------------------------|------|-----|---------------|-----|-----|-----|
| **General language composite** |      |     |               |     |     | 0.61|
| Cumulative L2 exposure         | 0.05 | 0.007 | [0.04, 0.07] | <0.001 | 0.47 |
| DLD                            | −1.02 | 0.154 | [−1.32, −0.71] | <0.001 | −0.58|
| Cumulative L2 exposure × DLD   | 0.01  | 0.015 | [−0.02, 0.04] | 0.726  | 0.02 |
| Non-verbal serial STM          | 0.04  | 0.111 | [−0.18, 0.26] | 0.724  | 0.04 |
| Cumulative L2 exposure × Non-verbal serial STM | 0.02  | 0.012 | [0.00, 0.04] | 0.121  | 0.14 |
| DLD × Non-verbal serial STM    | −0.36 | 0.223 | [−0.80, 0.08] | 0.110  | −0.17|
| Cumulative L2 exposure × DLD × Non-verbal serial STM | 0.05  | 0.024 | [0.00, 0.10] | 0.041  | 0.20 |
| **Expressive language composite** |      |     |               |     |     | 0.60|
| Cumulative L2 exposure         | 0.06  | 0.008 | [0.04, 0.07] | <0.001 | 0.50 |
| DLD                            | −1.04 | 0.166 | [−1.37, −0.71] | <0.001 | −0.57|
| Cumulative L2 exposure × DLD   | 0.01  | 0.016 | [−0.03, 0.04] | 0.744  | 0.02 |
| Non-verbal serial STM          | 0.07  | 0.120 | [−0.17, 0.31] | 0.568  | 0.06 |
| Cumulative L2 exposure × Non-verbal serial STM | 0.02  | 0.013 | [0.00, 0.05] | 0.097  | 0.16 |
| DLD × Non-verbal serial STM    | −0.25 | 0.240 | [−0.72, 0.23] | 0.302  | −0.11|
| Cumulative L2 exposure × DLD × Non-verbal serial STM | 0.05  | 0.026 | [0.00, 0.11] | 0.036  | 0.20 |
| **Receptive language composite** |      |     |               |     |     | 0.52|
| Cumulative L2 exposure         | 0.05  | 0.008 | [0.03, 0.07] | <0.001 | 0.46 |
| DLD                            | −0.82 | 0.174 | [−1.17, −0.48] | <0.001 | −0.47|
| Cumulative L2 exposure × DLD   | 0.01  | 0.017 | [−0.03, 0.04] | 0.754  | 0.02 |
| Non-verbal serial STM          | 0.02  | 0.126 | [−0.23, 0.26] | 0.904  | 0.01 |
| Cumulative L2 exposure × Non-verbal serial STM | 0.01  | 0.013 | [−0.01, 0.04] | 0.290  | 0.11 |
| DLD × Non-verbal serial STM    | −0.36 | 0.251 | [−0.86, 0.13] | 0.150  | −0.17|
| Cumulative L2 exposure × DLD × Non-verbal serial STM | 0.04  | 0.027 | [−0.01, 0.10] | 0.124  | 0.16 |
| **Language reasoning composite** |      |     |               |     |     | 0.63|
| Cumulative L2 exposure         | 0.05  | 0.008 | [0.03, 0.06] | <0.001 | 0.41 |
| DLD                            | −1.19 | 0.158 | [−1.50, −0.87] | <0.001 | −0.65|
| Cumulative L2 exposure × DLD   | 0.01  | 0.015 | [−0.03, 0.04] | 0.735  | 0.02 |
| Non-verbal serial STM          | 0.03  | 0.114 | [−0.19, 0.26] | 0.761  | 0.03 |
| Cumulative L2 exposure × Non-verbal serial STM | 0.02  | 0.012 | [0.00, 0.04] | 0.103  | 0.15 |
| DLD × Non-verbal serial STM    | −0.47 | 0.229 | [−0.92, −0.01] | 0.045  | −0.23|
| Cumulative L2 exposure × DLD × Non-verbal serial STM | 0.05  | 0.025 | [0.00, 0.10] | 0.035  | 0.21 |

Note: DLD, dummy variable, which before centring had 0 = typically developing children and 1 = children with developmental language disorder and after centring −0.49 and 0.51, respectively; STM, short-term memory. Also, four propensity score class dummy variables were included in the model to adjust TD and DLD group differences but reporting their coefficients is irrelevant.

*P*-values were calculated using small-sample degrees of freedom for multiple imputations (Reiter 2007; Van Ginkel and Kroonenberg 2014).

in DLD (Arslan et al. 2020, Vugs et al. 2013) but often failed to find impairments in simple visuo-spatial storage tasks (e.g., Arslan et al. 2020, Engel de Abreu et al. 2014). The present study did not aim to contrast verbal with visuo-spatial STM or WM. Instead, the interest was in STM for order. Based on previous research on the role of order memory in vocabulary acquisition (Cowan et al. 2017, Majerus and Boukebza 2013, Majerus et al. 2006b), it was hypothesized that the development of domain-general STM for temporal order would be atypical in DLD. For this purpose, two serial STM tasks were designed in the visual and auditory modality, respectively. A composite variable of these tasks (the average of their z-standardized scores) should control for modality-specific strategies, as in the composite, common variation is greater. In an earlier study (Lahti-Nuuttila et al. 2021), monolingual children with DLD were compared with their TD peers using this non-verbal serial STM composite variable. A pattern of effects was found suggesting that storing temporal order is difficult for children with DLD. Within the group
TABLE 6 Disentangling age and cumulative L2 exposure. Comparison of two sets of multiple regression models differing in the included interactions. The shared predictors of each of the language composites were standardized age, cumulative L2 exposure, group status, non-verbal serial short-term memory and their interactions. Model 1 also included interactions for age but not cumulative L2 exposure, whereas model 2 also included interactions for cumulative L2 exposure but not age.

| Model 1 | Model 2 |
|---------|---------|
| **β**  | **p**   | **R²**  | **β**  | **p**   | **R²**  |
| Age     | 0.57    | <0.001  | 0.49   | <0.001  |
| Cumulative L2 exposure | 0.25 | <0.001 | 0.31 | <0.001 |
| DLD     | -0.57   | <0.001  | -0.59  | <0.001  |
| Age × DLD | 0.10 | 0.146  | n.a.  | n.a.    |
| Cumulative L2 exposure × DLD | n.a. | n.a.  | -0.03 | 0.594   |
| Non-verbal serial STM | -0.12 | 0.183  | -0.12 | 0.178   |
| Age × Non-verbal serial STM | 0.28 | 0.008  | n.a.  | n.a.    |
| Cumulative L2 exposure × Non-verbal serial STM | n.a. | n.a.  | 0.19  | 0.012   |
| DLD × Non-verbal serial STM | -0.08 | 0.368  | -0.06 | 0.456   |
| Age × DLD × Non-verbal serial STM | 0.23 | 0.029  | n.a.  | n.a.    |
| Cumulative L2 exposure × DLD × Non-verbal serial STM | n.a. | n.a.  | 0.18  | 0.019   |
| **Expressive language composite** | 0.71 | 0.71 |
| Age     | 0.51    | <0.001  | 0.42   | <0.001  |
| Cumulative L2 exposure | 0.30 | <0.001 | 0.36 | <0.001 |
| DLD     | -0.54   | <0.001  | -0.58  | <0.001  |
| Age × DLD | 0.11 | 0.145  | n.a.  | n.a.    |
| Cumulative L2 exposure × DLD | n.a. | n.a.  | -0.02 | 0.703   |
| Non-verbal serial STM | -0.08 | 0.417  | -0.07 | 0.458   |
| Age × Non-verbal serial STM | 0.30 | 0.010  | n.a.  | n.a.    |
| Cumulative L2 exposure × Non-verbal serial STM | n.a. | n.a.  | 0.19  | 0.018   |
| DLD × Non-verbal serial STM | -0.02 | 0.807  | -0.02 | 0.842   |
| Age × DLD × Non-verbal serial STM | 0.23 | 0.054  | n.a.  | n.a.    |
| Cumulative L2 exposure × DLD × Non-verbal serial STM | n.a. | n.a.  | 0.19  | 0.024   |
| **Receptive language composite** | 0.69 | 0.69 |
| Age     | 0.60    | <0.001  | 0.51   | <0.001  |
| Cumulative L2 exposure | 0.23 | <0.001 | 0.29 | <0.001 |
| DLD     | -0.48   | <0.001  | -0.48  | <0.001  |
| Age × DLD | 0.15 | 0.071  | n.a.  | n.a.    |
| Cumulative L2 exposure × DLD | n.a. | n.a.  | -0.03 | 0.613   |
| Non-verbal serial STM | -0.12 | 0.226  | -0.15 | 0.139   |
| Age × Non-verbal serial STM | 0.23 | 0.048  | n.a.  | n.a.    |
| Cumulative L2 exposure × Non-verbal serial STM | n.a. | n.a.  | 0.15  | 0.073   |
| DLD × Non-verbal serial STM | -0.13 | 0.199  | -0.06 | 0.552   |
| Age × DLD × Non-verbal serial STM | 0.27 | 0.025  | n.a.  | n.a.    |
| Cumulative L2 exposure × DLD × Non-verbal serial STM | n.a. | n.a.  | 0.14  | 0.097   |
| **Language reasoning composite** | 0.77 | 0.78 |
| Age     | 0.55    | <0.001  | 0.48   | <0.001  |
| Cumulative L2 exposure | 0.20 | <0.001 | 0.24 | <0.001 |
| DLD     | -0.63   | <0.001  | -0.66  | <0.001  |
| Age × DLD | 0.04 | 0.566  | n.a.  | n.a.    |
| Cumulative L2 exposure × DLD | n.a. | n.a.  | -0.03 | 0.582   |
| Non-verbal serial STM | -0.14 | 0.108  | -0.12 | 0.158   |

(Continues)
with DLD, good serial STM appeared to support language acquisition.

In the current study, early sequential bilinguals with DLD were compared with bilingual TD children. It was hypothesized again that impairment of a domain-general capacity to represent order would be associated with DLD and atypical language development. Capacity for representing temporal order, reflected in serial STM performance, could have effects on language that might depend on either age, language exposure, or both. These variables could not be separated in the study of monolingual children. The present study was designed to examine the relationship between serial STM and DLD in children learning L2. Separate consideration of the effects of age and cumulative L2 exposure in bilingual children should make it possible to disentangle these two variables in the language acquisition of TD children and children with DLD.

To test the hypothesis that STM for order plays a special role in DLD, the development of non-verbal serial STM as a function of age was examined, on the one hand, and cumulative L2 exposure, on the other, comparing both TD children and children with DLD acquiring their second language. The results revealed that TD children’s serial STM capacity, as probed by the non-verbal order STM tasks, was greater than that of children with DLD as a function of both age and cumulative L2 exposure. The results replicated the previous findings in a sample of monolingual children (a sample that similarly consisted of TD children and children with DLD) with respect to serial STM development with age (Lahti-Nuuttila et al. 2021). These earlier results suggested that a domain-general mechanism for presenting temporal order develops atypically in DLD. However, there is no a priori reason to expect domain-general memory for order to develop with exposure to a second language. In the present study, evidence for a relationship between cumulative L2 exposure and STM was also found. The most likely explanation for the result lies in the moderate correlations between age and cumulative L2 exposure in this sample. This interpretation is supported by the finding that the moderation effect for L2 exposure on the effect of DLD on STM was no longer significant when age was included in the same regression model (table 3, model 3). In addition to age, the amount of experience with Finnish daycare pedagogics covaries with L2 exposure in the present data set. Potentially these structured but unspecified organized activities can also contribute to STM development, showing up in the effect of the operationalization of cumulative language exposure in this study.

To test the second hypothesis that serial STM is related to language development in DLD, moderation by non-verbal serial STM of the effects of age and cumulative L2 exposure on different language composites in the two groups of children was studied. Both age and L2 exposure had stronger effects in TD children compared with children with DLD, reflecting faster language acquisition with both age and exposure in TD. The moderation of the effect of age by serial STM was found only in the children with DLD, especially robustly in expressive and receptive language. For them, better non-verbal serial STM was associated with greater improvements in language measures with increasing age. Similar, but perhaps smaller, moderation effects were also found for cumulative L2 exposure on the expressive language and the language reasoning composites, whereas receptive language only showed a non-significant trend. The moderation effects on measures of language development with age suggest that memory for serial order could play a role in language acquisition in DLD. There was a very similar pattern of serial STM moderation of the relationship between exposure and the language composites. The differences between effects on different language components have to be treated with caution, as the psychometric test items that account for individual variation are different at different ages (moving from word level to sentence level in some tests) and possibly differently sensitive to the amount of language exposure. Future studies with experimentally constructed tasks will be able to probe in detail the different aspects of language development in relation to serial STM.
When moderation effects were studied for one of the variables (age or cumulative exposure) while controlling the main effect of the other variable, the moderation of age was found to be statistically significant for receptive language. However, for expressive language and language reasoning, it was the moderation of cumulative exposure that was significant. Here the effect sizes do not differ very much, and in a complex model with moderate sample size, the interpretation must be cautious because age and cumulative exposure covary, so this result could be a statistical artefact. However, the result could also truly indicate that with cumulative exposure controlled, serial STM moderation is somewhat different. This needs to be tested with a targeted study design in the future.

As the pattern of serial STM development seems similar for mono- and bilingual children with DLD, the present study added support for the hypothesis that domain-general serial STM development between the ages of four and seven years is impaired in DLD. This study also replicated the findings from the study of monolingual children (Lahti-Nuuttila et al. 2021) that serial STM moderates language development in children with DLD in this age range but not in TD children. These findings can speculatively be explained by assuming that impaired STM for order is part of the clinical picture of DLD. Further, when serial STM is in the impaired range, it tends to be associated with slower than typical language development. In TD children, individual differences in domain-general serial STM are in a range that does not appear to be related to language development. The found moderation effect in children with DLD could also suggest that effective non-verbal serial STM could be used as a compensation mechanism in atypical language development. Since this study was cross-sectional, causal (possibly reciprocal) relations between serial STM and language need to be studied further with a longitudinal design to rule out the possibility that it is the language difficulties that affect non-verbal serial STM.

Several aspects of this study are problematic for the suggested interpretation. First, the range of PIQ values in the group with DLD included values between 70 and 85. In the original protocol, the plan was to treat these children separately. However, as the criteria for DLD were revised with CATALISE (Bishop et al. 2017) and as the children with DLD and PIQ of 70–84 formed roughly a third of the clinical sample, it was judged that excluding them would lead to a misrepresentation of the clinical DLD population in Finland. In studies of DLD, the non-verbal ability of the DLD group is often somewhat lower than that of the TD control group (e.g., Cowan et al. 2017). This may have been accentuated in the present study because the PIQ score in WPPSI-III is partly based on the subtest Picture Concepts, which has also been associated with the verbal ability (Peyre et al. 2016, Saar et al. 2018). This could have led to an underestimate of non-verbal reasoning skills, especially in children with DLD tested in their L2. Also, among the bilingual TD children, this possibly led to increased exclusions from the control group as in the additional subsequent analyses we noticed that bilingual TD children had more often low standard scores specifically on this subtest compared with the other two PIQ components (Block Design and Matrix Reasoning).

As the PIQ estimates were low for some of the children with DLD, it is possible that some of them may later be given a different diagnosis, for example, general learning or intellectual disability. However, in the present study, these children did not show atypical adaptive reasoning capacity in their daily lives as reported by their parents or as observed by the multidisciplinary team in the Audiophoniciatric Ward. Furthermore, in the present study, propensity scores were used to balance the two groups for differences in non-verbal cognition.

A second limitation of the study is the reliability of the serial STM measure among the younger children. Although this study showed a significant difference between the two groups of children in STM improvement with age, the question remains if, with better measures, the difference could be found already at younger ages. One possibility could be to increase the number of trials in the STM tasks to improve sensitivity to impairment among the youngest children. This would not necessarily make the task much longer as the increase could be restricted to short series lengths that are close to the youngest children’s serial STM capacity limit.

A third consideration concerns the construct of serial STM or STM for order. Other researchers have reported impairments in sustained attention related to DLD (Boerma et al. 2017, Ebert and Kohnert 2011, Ebert et al. 2019, Finneran et al. 2009) and sustained attention may be one of the cognitive components enabling performance in serial STM tasks. Related to this, some children, especially in the DLD group, might later show even symptoms of comorbid attention deficit hyperactivity disorder (ADHD) as it is two to three times more likely for children with language impairment to have ADHD than for TD children (Mueller and Tomblin 2012). Unfortunately, the comorbidity of DLD and ADHD could not be taken into account in the current study. According to the Finnish edition of ICD-10 (WHO 2010), ADHD is difficult to detect in children before school age, that is, under the age of seven years, due to the wide normal variation, and the diagnosis should be made only in extreme cases. The studied children had not started school, were young to be evaluated for ADHD and did not present extreme characteristics in the structured clinical examination and interview including background information and questionnaires. In addition to the possible role of attention in the STM tasks (Hakim et al. 2020),
clinical or subclinical deficits in attention are likely to play a role also in language development. A review of cognitive skills supporting language comprehension and production in adults (Federmeier et al. 2020) reveals how entangled domain-general cognitive processes, such as attention-dependent executive processes, sustained attention, and the ability to control information flow over time, are with language. Future targeted research is needed to reveal what role sustained or other attention plays in both representing serial order in STM and language development.

Another question for future research concerns the set of specific processes in STM that operate in the kind of STM task used here. To give just one example, it could be that some children have better strategies for naming the stimuli and perhaps subvocally rehearsing them also in non-verbal serial STM tasks. The stimuli in this study were designed to minimize naming in young children (Gathercole et al. 1994), but variation in strategic approaches to the task cannot be totally ruled out. Questions about the children’s use of verbal and other strategies remain to be resolved in more targeted research.

Also, an interesting subject for future research is the possible effect of speech and language therapy. In the DLD group, some children had had speech and language therapy but, in this study, this could not be taken into account because of the small sample size. To some extent, L2 therapy was included in the cumulative exposure, but certainly a longitudinal intervention study would be more informative.

Finally, some background variables could introduce confounds into the data. A potential problem might be the unequal distribution of L1s in the DLD and TD groups (see table S1 in the additional supporting information). Estonian is the only L1 in the sample that is closely related to Finnish. There were seven more children with Estonian as L1 in the TD group than in the DLD group. Although the number of Estonian L1 children was small, to control the possible effect of Estonian as L1, analyses with it as a binary dummy variable were run. In these analyses, the effects remained very similar to the ones reported. Another interesting background variable whose impact should be studied in the future is socioeconomic status. Inclusion of mother’s education in years in the preliminary analyses did not change the central results in the present data set.

CONCLUSIONS

This study was designed to explore whether deficits in non-verbal STM for order are associated with bilingual DLD. A sample of 4–6-year-old bilingual children with DLD and TD children was studied, assessed in their second language. The serial STM of children with DLD was found to be poorer and to show less improvement with age than that of TD children. Furthermore, the improvement of language performance as a function of age or L2 exposure, detected by composite measures of receptive language, expressive language, and language reasoning, was moderated by STM in children with DLD but not in TD children of this age range. We conclude that STM for order, measured by simple non-verbal game-like tasks, can be helpful in comprehending and planning interventions for DLD in young children learning their second language.

ACKNOWLEDGEMENTS

The authors are grateful to the participating children and their families and to the speech language therapists, psychologists, phoniatricians, nurses and other personnel at the Department of Phoniatrics, University of Helsinki, and Helsinki University Hospital, as well as to the participating kindergartens and their personnel. For their invaluable contributions to this work, we thank Miika Leminen MSc, MPsych; software developer Iida Porokuokka MSc; Erkki Vilkman MD, PhD; and Ahmed Geneid MD, PhD. This study is part of a larger research project, the Helsinki longitudinal SLI study (Laasonen et al. 2018) and its cognitive subproject.

DECLARATION OF INTEREST

The authors declare that they have no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data are available on request due to privacy/ethical restrictions. Requests to access the dataset should be directed to Marja Laasonen.

ORCID

Pekka Lahti-Nuuttila https://orcid.org/0000-0001-5463-1738
Marja Laasonen https://orcid.org/0000-0002-4628-4251
Sini Smolander https://orcid.org/0000-0003-0517-0298
Sari Kunnari https://orcid.org/0000-0001-5290-4851
Eva Arkkila https://orcid.org/0000-0003-0067-3216
Elisabet Service https://orcid.org/0000-0002-7698-1189

NOTE

1 At most there were 12 regressor terms in the analyses. A priori power analyses for adequate sample size were run as described by Laasonen et al. (2018).

REFERENCES

Archibald, L.M.D. (2017) Working memory and language learning: a review. Child Language Teaching & Therapy, 33, 5–17. https://doi.org/10.1177/0261561316654206
Archibald, L.M.D. & Gathercole, S.E. (2006a) Short-term and working memory in specific language impairment. International
Journal of Language & Communication Disorders, 41, 675–693. 
https://doi.org/10.1080/13682820500442602

Archibald, L.M.D. & Gathercole, S.E. (2006b) Visuospatial immediate memory in specific language impairment. Journal of Speech, Language, and Hearing Research, 49, 265–277. 
https://doi.org/10.1044/1092-4388(2006/022)

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

Attout, L., Grégoire, C. & Majerus, S. (2020) How robust is the link between working memory for serial order and lexical skills in children? Cognitive Development, 53, 100854. 
https://doi.org/10.1016/j.cogdev.2020.100854

Austin, P.C. (2011) An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behavioral Research, 46, 399–424. 
https://doi.org/10.1080/00273611.2011.586786

Baddeley, A. (2003) Working memory and language: an overview. Journal of Communication Disorders, 36, 189–208. 
https://doi.org/10.1016/S0021-9924(03)00019-4

Baddeley, A., Gathercole, S. & Papagno, C. (1998) The phonological loop as a language learning device. Psychological Review, 105, 158–173. 
https://doi.org/10.1037/0033-295X.105.1.158

Bishop, D.V.M., 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, 1068–1080. 
https://doi.org/10.1111/jcpp.12721

Boerma, T., Chiat, S., Leseman, P., Timmermeister, M., Wijnen, F. & Blom, E. (2015) A quasi-universal nonword repetition task as a diagnostic tool for bilingual children learning Dutch as a second language. Journal of Speech, Language, and Hearing Research, 58, 1747–1760. 
https://doi.org/10.1044/2015_JSLHR-L-15-0058

Boerma, T., Leseman, P., Wijnen, F. & Blom, E. (2017) Language proficiency and sustained attention in monolingual and bilingual children with and without language impairment. Frontiers in Psychology, 8, 1241. 
https://doi.org/10.3389/fpsyg.2017.01241

Clark, G.M. & Lum, J.A.G. (2017) Procedural memory and speed of grammatical processing: comparison between typically developing children and language impaired children. Research in Developmental Disabilities, 71, 237–247. 
https://doi.org/10.1016/j.ridd.2017.10.015

Cowan, N., Hogan, T.P., Alt, M., Green, S., Cabbage, K.L., Brinkley, S. & Gray, S. (2017) Short-term memory in childhood dyslexia: deficient serial order in multiple modalities. Dyslexia, 23, 209–233. 
https://doi.org/10.1002/dys.1557

Ebert, K.D. & Kohnert, K. (2011) Sustained attention in children with primary language impairment: a meta-analysis. Journal of Speech, Language & Hearing Research, 54, 1372–1384. 
https://doi.org/10.1044/1092-4388(2011/10-0231)

Ebert, K.D., Rak, D., Slawny, C.M. & Fogg, L. (2019) Attention in bilingual children with developmental language disorder. Journal of Speech, Language, and Hearing Research, 62, 979–992. 
https://doi.org/10.1044/2018_JSLHR-L-18-0221

Edwards, S., Fletcher, P., Garman, M., Hughes, A., Letts, C. & Sinka, I. (1997) Reynell Developmental Language Scales III. Windsor, UK: Nfer-Nelson [Translation and standardisation of the Finnish version by Kortesmaa, M., Heimonen, K., Merikoski, H., Warma, M.-L. & Varpela, V., (2001) Reynellin kielellisen kehityksen testi. Helsinki: Psykologien Kustannus Oy]

Engel De Abreu, P.M.J., Cruz-Santos, A. & Puglisi, M.L. (2014) Specific language impairment in language-minority children from low-income families. International Journal of Language & Communication Disorders, 49, 736–747. 
https://doi.org/10.1111/1460-6984.12107

Federmeier, K.D., Jongman, S.R. & Szewczyk, J.M. (2020) Examining the role of general cognitive skills in language processing: a window into complex cognition. Current Directions in Psychological Science, 29, 575–582. 
https://doi.org/10.1177/0963721420964095

Finneran, D.A., Francis, A.L. & Leonard, L.B. (2009) Sustained attention in children with specific language impairment (SLI). Journal of Speech, Language, and Hearing Research, 52, 915–929. 
https://doi.org/10.1044/2019-4388(2009/07-0053)

Gallinat, E. & Spaulding, T.J. (2014) Differences in the performance of children with specific language impairment and their typically developing peers on nonverbal cognitive tests: a meta-analysis. Journal of Speech, Language, and Hearing Research, 57, 1363–1382. 
https://doi.org/10.1044/2014_JSLHR-L-12-0363

Gathercole, S.E., Adams, A.-M. & Hitch, G. (1994) Do young children rehearse? An individual-differences analysis. Memory & Cognition, 22, 201–207. 
https://doi.org/10.3758/bf03208891

Gathercole, S.E. & Baddeley, A.D. (1990) Phonological memory deficits in language disordered children: is there a causal connection? Journal of Memory and Language, 29, 336–360. 
https://doi.org/10.1016/S0749-596X(90)80004-1

Gathercole, S.E. & Baddeley, A.D. (1993) Phonological working memory: a critical building block for reading development and vocabulary acquisition? European Journal of Psychology of Education, 8, 259–272. 
https://doi.org/10.1007/BF03374081

Girbau, D. & Schwartz, R.G. (2008) Phonological working memory in Spanish–English bilingual children with and without specific language impairment. Journal of Communication Disorders, 41, 124–145. 
https://doi.org/10.1016/j.jcomdis.2007.07.001

Hakim, N., Debettencourt, M.T., Awh, E. & Vogel, E.K. (2020) Attention fluctuations impact ongoing maintenance of information in working memory. Psychonomic Bulletin & Review, 27, 1269–1278. 
https://doi.org/10.3758/s13423-020-01790-z

Hartley, T., Hurlstone, M.J. & Hitch, G.J. (2016) Effects of rhythm on memory for spoken sequences: a model and tests of its stimulus-driven mechanism. Cognitive Psychology, 87, 135–178. 
https://doi.org/10.1016/j.cogpsych.2016.05.001

Hayes, A.F. (2018) Introduction to mediation, moderation, and conditional process analysis: a regression-based approach. 2nd edition. New York, NY: Guilford Press.

Hayes, A.F. & Rockwood, N.I. (2017) Regression-based statistical mediation and moderation analysis in clinical research: observations, recommendations, and implementation. Behaviour Research and Therapy, 98, 39–57. 
https://doi.org/10.1016/j.brat.2016.11.001

Henry, L.A. & Botting, N. (2017) Working memory and developmental language impairments. Child Language Teaching & Therapy, 33, 19–32. 
https://doi.org/10.1177/0265690116655378

Hsu, H.J. & Bishop, D.V. (2011) Grammatical Difficulties in Children with Specific Language Impairment: is Learning
Deficient? Human Development, 53, 264–277. https://doi.org/10.1159/000321289

Hurlstone, M.J. (2019) Functional similarities and differences between the coding of positional information in verbal and spatial short-term order memory. Memory (Hove, England), 27, 147–162. https://doi.org/10.1080/09658211.2018.1495235

Hurlstone, M.J. & Hitch, G.J. (2018) How is the serial order of a visual sequence represented? Insights from transposition latencies. Journal of Experimental Psychology: Learning, Memory, and Cognition, 44, 167–192. https://doi.org/10.1037/xlm0000440

Hurlstone, M.J., Hitch, G.J. & Baddeley, A.D. (2014) Memory for serial order across domains: an overview of the literature and directions for future research. Psychological Bulletin, 140, 339–373. https://doi.org/10.1037/a0034221

Kaplan, E., Goodglass, H., Weintraub, S. & Segal, O. (1983) Boston Naming Test. Philadelphia, PA: Lea & Febiger

Kohnert, K. (2010) Bilingual children with primary language impairment: issues, evidence and implications for clinical actions. Journal of Communication Disorders, 43, 456–473. https://doi.org/10.1016/j.jcomdis.2010.02.002

Kohnert, K., Windsor, J. & Ebert, K.D. (2009) Primary or “specific” language impairment and children learning a second language. Brain and Language, 109, 101–111. https://doi.org/10.1016/j.bandl.2008.01.009

Korkman, M., Kirk, U. & Kemp, S.L. (2008) NEPSY-II: Lastenneuro psykologinen tutkimus. Helsinki, Finland: Psykologien Kustannus.

Laasonen, M., Smolander, S., Lahti-Nuuttila, P., Leminen, M., Lajunen, H.-R., Heinonen, K., et al. (2018) Understanding developmental language disorder—the Helsinki longitudinal SLI study (HeSLI): a study protocol. BMC Psychology, 6, 24. https://doi.org/10.1186/s40359-018-0222-7

Lahti-Nuuttila, P., Service, E., Smolander, S., Kunnari, S., Arkkila, E. & Laasonen, M. (2021) Short-Term Memory for Serial Order Moderates Aspects of Language Acquisition in Children With Developmental Language Disorder: findings From the HeSLI Study. Frontiers in Psychology, 12. https://doi.org/10.3389/fpsyg.2021.608069

Leclercq, A.-L. & Majerus, S. (2010) Serial-order short-term memory predicts vocabulary development: evidence from a longitudinal study. Developmental Psychology, 46, 417–427. https://doi.org/10.1037/a0018540

Leonard, L.B., Weismer, S.E., Miller, C.A., Francis, D.J., Tomblin, J.B. & Kail, R.V. (2007) Speed of processing, working memory, and language impairment in children. Journal of Speech, Language, and Hearing Research, 50, 408–428. https://doi.org/10.1044/1092-4388(2007/029

Majerus, S. & Boukeba, C. (2015) Short-term memory for serial order supports vocabulary development: new evidence from a novel word learning paradigm. Journal of Experimental Child Psychology, 116, 811–828. https://doi.org/10.1016/j.jecp.2013.07.014

Majerus, S., Poncelet, M., Elsen, B. & van der Linden, M. (2006a) Exploring the relationship between new word learning and short-term memory for serial order recall, item recall, and item recognition. European Journal of Cognitive Psychology, 18, 848–873. https://doi.org/10.1080/09541440500446476

Majerus, S., Poncelet, M., Greffe, C. & van der Linden, M. (2006b) 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. https://doi.org/10.1016/j.jecp.2005.07.005

Martin, N.A. & Brownell, R. (2010) Receptive One-Word Picture Vocabulary Test-4 (ROWPVT-4). Novato, CA: Academic Therapy Publications. [Translation and standardisation of the Finnish version by Kunnari, S. & Välimaa, T., (in validation)]

Martin, N.A. & Brownell, R. (2011) Expressive One-Word Picture Vocabulary Test-4 (EOWPVT-4). Novato, CA: Academic Therapy Publications. [Translation and standardisation of the Finnish version by Kunnari, S. & Välimaa, T., (in validation)]

Montgomery, J.W., Magimairaj, B.M. & Finney, M.C. (2010) Working memory and specific language impairment: an update on the relation and perspectives on assessment and treatment. American Journal of Speech–Language Pathology, 19, 78–94. https://doi.org/10.1044/1058-0360(2009/09-0028

Mueller, K.L. & Tomblin, J.B. (2012) Examining the comorbidity of language impairment and attention-deficit/hyperactivity disorder. Topics in Language Disorders, 32, 228–246. https://doi.org/10.1097/TLD.0b013e318262010d

Ordonez Magro, L., Attout, L., Majerus, S. & Szmalec, A. (2018) Short- and long-term memory determinants of novel word form learning. Cognitive Development, 47, 146–157. https://doi.org/10.1016/j.cogdev.2018.06.002

Paradis, J. (2011) Individual differences in child English second language acquisition: comparing child-internal and child-external factors. Linguistic Approaches to Bilingualism, 1, 213–237. https://doi.org/10.1075/lab.1.3.01par

Peyre, H., Bernard, J.Y., Hoertel, N., Forhan, A., Charles, M.-A., De Agostini, M., Heude, B., Ramsus, F. & the EDEN Mother-Child Cohort Study Group. (2016) Differential effects of factors influencing cognitive development at the age of 5-to-6 years. Cognitive Development, 40, 152–162. https://doi.org/10.1016/j.cogdev.2016.10.001

Reiter, J.P. (2007) Small-sample degrees of freedom for multicomponent significance tests with multiple imputation for missing data. Biometrika, 94, 502–508. https://doi.org/10.1093/biomet/asm028

Rosenbaum, P.R. & Rubin, D.B. (1983) The central role of the propensity score in observational studies for causal effects. Biometrika, 70, 41–55. https://doi.org/10.2307/2335942

Saar, V., Levänen, S. & Komulainen, E. (2018) Cognitive profiles of Finnish preschool children with expressive and receptive language impairment. Journal of Speech, Language, and Hearing Research, 61, 386–397. https://doi.org/10.1044/2017_JSLHR-L-16-0365

Schafer, J.L. & Kang, J. (2008) Average causal effects from non-randomized studies: a practical guide and simulated example. Psychological Methods, 13, 279–313. https://doi.org/10.1037/a0014268

Smolander, S., Laasonen, M., Arkkila, E., Lahti-Nuuttila, P. & Kunnari, S. (2021) L2 vocabulary acquisition of early sequentially bilingual children with TD and DLD affected differently by exposure and age of onset. International Journal of Language & Communication Disorders, Advance online publication. 56, 72–89. https://doi.org/10.1111/1460-6984.12583

Ullman, M.T. & Pierpont, E.I. (2005) Specific language impairment is not specific to language: the procedural deficit hypothesis. Cortex, 41, 399–433. https://doi.org/10.1016/s0010-9452(08)70276-4

van Ginkel, J.R. & Kroonenberg, P.M. (2014) Analysisofvarianceofrelationsbetweenvocabularydevelopmentandverbalshort-termorderanditeminformation. [TranslationandstandardisationoftheFinnishtestbyKunnari,S.&Välimaa,T.,(invalidation)]

van Ginkel, J.R. & Kroonenberg, P.M. (2014) [Translation and standardisation of the Finnish version by Kunnari, S. & Välimaa, T., (in validation)]
Verhagen, J. & Leseman, P. (2016) How do verbal short-term memory and working memory relate to the acquisition of vocabulary and grammar? A comparison between first and second language learners. *Journal of Experimental Child Psychology*, 141, 65–82. [https://doi.org/10.1016/j.jecp.2015.06.015](https://doi.org/10.1016/j.jecp.2015.06.015)

Vugs, B., Cuperus, J., Hendriks, M. & Verhoeven, L. (2013) Visuospatial working memory in specific language impairment: a meta-analysis. *Research in Developmental Disabilities*, 34, 2586–2597. [https://doi.org/10.1016/j.ridd.2013.05.014](https://doi.org/10.1016/j.ridd.2013.05.014)

Wechsler, D. (2009) *WPPSI-III - Wechsler preschool and primary scale of intelligence*. 3rd Edition, Helsinki: Psykologien Kustannus.

Windsor, J., Kohnert, K., Lobitz, K.F. & Pham, G.T. (2010) Cross-language nonword repetition by bilingual and monolingual children. *American Journal of Speech–Language Pathology*, 19, 298–310. [https://doi.org/10.1044/1058-0360(2010/09-0064)](https://doi.org/10.1044/1058-0360(2010/09-0064))

World Health Organization (WHO) (2010) *ICD-10: International statistical classification of diseases and related health problems: 10th revision*. 3rd edition Geneva: World Health Organization. [Finnish version: Terveyden ja hyvinvoinnin laitos (THL), (2011) Tautiluokitus ICD-10. 3. painos. Helsinki: Terveyden ja hyvinvoinnin laitos]

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Lahti-Nuuttila P, Laasonen M, Smolander S, Kunnari S, Arkkila E, Service E. Language acquisition of early sequentially bilingual children is moderated by short-term memory for order in developmental language disorder: Findings from the HeISLI study. *International Journal of Language & Communication Disorders*. 2021;56:907–926. [https://doi.org/10.1111/1460-6984.12635](https://doi.org/10.1111/1460-6984.12635)