



# Failure to consolidate statistical learning in developmental dyslexia

Ranin Ballan<sup>1,2</sup> · Simon J. Durrant<sup>3</sup> · Dara S. Manoach<sup>4,5,6</sup> · Yafit Gabay<sup>1,2</sup>

Accepted: 7 August 2022 / Published online: 11 October 2022  
© The Psychonomic Society, Inc. 2022

## Abstract

Statistical learning (SL), the ability to pick up patterns in sensory input, serves as one of the building blocks of language acquisition. Although SL has been studied extensively in developmental dyslexia (DD), much less is known about the way SL evolves over time. The handful of studies examining this question were all limited to the acquisition of motor sequential knowledge or highly learned segmented linguistic units. Here we examined memory consolidation of statistical regularities in adults with DD and typically developed (TD) readers by using auditory SL requiring the segmentation of units from continuous input, which represents one of the earliest learning challenges in language acquisition. DD and TD groups were exposed to tones in a probabilistically determined sequential structure varying in difficulty and subsequently tested for recognition of novel short sequences that adhered to this statistical pattern in immediate and delayed-recall sessions separated by a night of sleep. SL performance of the DD group at the easy and hard difficulty levels was poorer than that of the TD group in the immediate-recall session. Importantly, DD participants showed a significant overnight deterioration in SL performance at the medium difficulty level compared to TD, who instead showed overnight stabilization of the learned information. These findings imply that SL difficulties in DD may arise not only from impaired initial learning but also due to a failure to consolidate statistically structured information into long-term memory. We hypothesize that these deficits disrupt the typical course of language acquisition in those with DD.

**Keywords** Developmental dyslexia · Memory consolidation · Segmentation · Statistical learning · Striatum

## Introduction

Our world is full of rich sensory patterns and regularities that can be exploited to guide behavior through a process of statistical learning (SL) (Conway, 2020). SL is believed to

represent an unsupervised form of learning in which learners implicitly extract structure from the environment. Similar to other forms of implicit learning, SL occurs in the absence of conscious awareness of patterns embedded in the input (Turk-Browne et al., 2009), and even while participants' attention is distracted by performing a concurrent secondary task (Horváth et al., 2020; Saffran et al., 1997; but see Toro et al., 2005). These characteristics situate SL as a form of learning that is more aligned with the procedural memory system rather than with the declarative memory system (Sawi & Rueckl, 2019), though this may depend on the nature of the training experience (Reber et al., 2003).

SL is believed to play a significant role in the acquisition of spoken and written linguistic skills (Arciuli & Simpson, 2012; Giustolisi & Emmorey, 2018; Qi et al., 2018; Spencer et al., 2015), and thus prompted much interest in the research field of developmental dyslexia (DD) (Lee et al., 2022; Saffran, 2018; Schmalz et al., 2017; Singh & Conway, 2021). DD is a specific and significant deficiency in the development of reading skills that is not solely accounted for by sensory impairments, neurological disorders, or inadequate schooling.

✉ Ranin Ballan  
raninb92@gmail.com

✉ Yafit Gabay  
ygabay@edu.haifa.ac.il

<sup>1</sup> Department of Special Education, University of Haifa, 31905 Haifa, Israel

<sup>2</sup> Edmond J. Safra Brain Research Center for the Study of Learning Disabilities, University of Haifa, Haifa, Israel

<sup>3</sup> Lincoln Sleep Research Centre and School of Psychology, University of Lincoln, Lincoln, UK

<sup>4</sup> Harvard Medical School, Boston, MA, USA

<sup>5</sup> Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA

<sup>6</sup> Athinoula A. Martinos Center for Biomedical Imaging, Charlestown, MA, USA

Traditionally, DD has been suggested to arise from phonological impairments (Snowling, 2001), but mounting evidence points to broader impairments (for a review see, Démonet et al., 2004), including statistical learning deficits. There is a growing body of evidence gleaned from independent investigations suggesting that people with DD are less capable of tracking patterns in their environment compared to typical readers (e.g., Gabay, 2021; Gabay et al., 2012b; Howard Jr. et al., 2006; Kahta & Schiff, 2019; Lum et al., 2013; Stoodley et al., 2006; Vicari et al., 2005). Yet other studies point to preserved statistical learning in DD (Kelly et al., 2002; Rüsseler et al., 2006).

SL abilities of individuals with DD have been examined across a variety of different experimental paradigms that are considered “SL” (for reviews see, Arciuli & Conway, 2018; Bogaerts et al., 2021; Singh & Conway, 2021). These different paradigms likely tap different cognitive processes, making it difficult to conclude about the nature of SL impairments in DD (Bogaerts et al., 2021). One paradigm that mimics language learning in real life (the ability to split a speech stream into units) is the segmentation task, in which listeners extract knowledge about transitional probabilities from continuous input that are higher within than across words (Saffran et al., 1996). These types of paradigms reveal a close relationship between SL abilities and spoken and written linguistic skills (Arciuli & Simpson, 2012; Giustolisi & Emmorey, 2018; Qi et al., 2018; Spencer et al., 2015). Consistently, studies that have used this paradigm discovered that adults (Gabay et al., 2015; Sigurdardottir et al., 2017) and children with DD (Singh et al., 2018; Tong et al., 2020; Tong et al., 2019; but see van Witteloostuijn et al., 2019) are less capable of incidentally extracting statistical regularities than neurotypicals.

## Memory consolidation

Although SL has been studied extensively in DD research, less attention has been paid to how this type of learning evolves over time. In addition to performance gains that can be observed within the training session (online learning, fast learning), additional gains are believed to occur in the absence of any additional training (offline learning, slow learning). These delayed gains are believed to reflect memory consolidation, that is, the process by which memories become less susceptible to interference and are honed to represent new knowledge (Dudai et al., 2015). Memory consolidation takes many forms (Walker, 2005). Stabilization refers to the strengthening of a memory trace after its acquisition (Robertson et al., 2004). Evidence for stabilization can be found in the loss of acquired knowledge if a person immediately attempts to acquire a similar task but not if there is an interval of time between learning

the first and second tasks (Brashers-Krug et al., 1996). Further behavioral improvement can be seen in the additional consolidation-based enhancement stage. During this stage, in the absence of any further rehearsal or experience, gains in performance may take place (Stickgold et al., 2000).

Memory consolidation processes have been observed across a variety of domains (Ben-Zion et al., 2022; Censor & Sagi, 2008; Davis et al., 2009; Earle & Myers, 2015; Earle & Myers, 2013; Janacek & Nemeth, 2012; Saltzman & Myers, 2021; Stickgold, 2005), including SL (Durrant et al., 2011; Durrant et al., 2013, 2016). For example, in the study conducted by Durrant et al. (2011), offline gains of SL were observed after a sleep interval. Other studies showed maintenance of SL following a sleep interval (Baran et al., 2018; Kim et al., 2009). Furthermore, sleep-dependent consolidation of auditory SL was accompanied by increased activity in the striatum (Durrant et al., 2013). In this regard, empirical and theoretical research suggests that DD is associated with a selective impairment in the striatal-based memory system (Fawcett & Nicolson, 2019; Nicolson & Fawcett, 2011; Ullman, 2004; Ullman et al., 2020). Therefore, based on evidence suggesting the involvement of the striatum in both fast and slow learning of statistical regularities (Durrant et al., 2013; Karuza et al., 2013), one could speculate that SL impairments in DD are likely to be evident not only during the initial stages of learning but also during later stages that involve memory consolidation of statistically structured information.

Notably, most studies investigating SL in DD and striatal-based learning in general examined learning in one session, thus disregarding later stages involved in the process of memory formation such as consolidation. Although there is a growing literature on memory consolidation in DD across different types of learning (Hollander & Adi-Japha, 2021; Reda et al., 2021; Smith et al., 2018) including SL (Hedenius et al., 2021), most studies examining SL in DD over time have concentrated on the motor domain and have used tasks not clearly related to language acquisition (Gabay et al., 2012a; Hedenius et al., 2013; Hedenius et al., 2020; Nicolson et al., 2010). A handful of studies examined consolidation of the structural properties of already-learned linguistic units, as in the case of artificial grammar learning (Inácio et al., 2018) or Hebb learning tasks examined across native-language syllables (Bogaerts et al., 2015). Notably, acquisition of language involves earlier learning challenges in which functional units must first be segmented from a continuous sound stream without a priori knowledge of the temporal time window that characterizes these units (Vihman, 2009), and the success of this learning is significant for later learning challenges in language acquisition (Finn & Hudson Kam, 2015).

## Experiment 1

In the present study, our aim was to examine segmentation-based SL in DD over time, which is considered one of the earliest learning challenges in language acquisition and has been shown to be closely related to language acquisition. We examined SL using a paradigm in which listeners are required to implicitly learn the contingent probabilities within a stream of input (Durrant et al., 2011). The paradigm we used is different from Saffran's segmentation paradigm in two important aspects. First, in the current task, zero/first/second-order contingent probabilities were controlled (depending on the task version), while the Saffran triplet approach always has a mixture of first- and second-order information that is never precisely quantified. This helps examine the interaction between SL and task difficulty in DD. Second, the current task is built in such a way that there are no explicitly created triplets, to avoid chunking and to reduce the involvement of declarative memory processing. Finally, this paradigm has been shown to engage the striatal-based memory system in neurotypicals during a consolidation phase (Durrant et al., 2013) and as such could help clarify the relationship between SL impairments and a striatal-learning deficit presumed to be associated with DD (Bogaerts et al., 2020).

## Methods

### Participants

Forty-two university students (21 with developmental dyslexia and 21 controls) took part in the study. All participants were native Hebrew speakers, had no history of neurological and/or psychiatric disorders (according to the American Psychiatric Association, 2013), and had normal or corrected-to-normal vision and normal hearing. The inclusion criteria for the dyslexia group were (1) a formal diagnosis from a licensed clinician; (2) the absence of a formal diagnosis of attention deficit hyperactivity disorder (ADHD) or a specific language impairment; (3) a score below a 1SD local norm cut-off for phonological decoding (Weiss et al., 2015); (4) IQ estimate within the normal range (Raven score > 10th percentile). Based on these criteria, one participant with DD was excluded from the final sample. The typically developing (TD) readers group was composed of individuals with no history of learning disabilities who exhibited no difficulties in reading [e.g., were above the reading cutoff (non-word reading)] and were at the same level of cognitive skills (assessed by the Raven test) as the DD group. The Institutional Review Board of

the University of Haifa approved the study, which was conducted in accordance with the Declaration of Helsinki, with written informed consent provided by all participants. Participants received compensation of NIS 120 (approximately \$37) for participating in the study.

Participants underwent a series of tests to evaluate cognitive and linguistic abilities, assessed by the Raven's Standard Progressive Matrices (Raven & Court, 1992), verbal short-term memory (as measured by the Digit Span subtest from the Wechsler Adult Intelligence Scale; Wechsler, 1997), rapid automatized naming skills (RAN; Breznitz & Misra, 2003), reading skills of words and non-words (Shatil, 1995a), and phonological processing (Breznitz & Misra, 2003). Participants also completed measures of sleep quality (Pittsburgh Sleep Quality Index (PSQI); Buysse et al., 1989) and alertness (Stanford Sleepiness Scale (SSS); Hoddes et al., 1973). Details of these tasks are presented in Table 1.

The groups did not differ in age and IQ estimate but compared to the TD group the DD group displayed a profile of reading disability compatible with the symptomatology of DD (Table 2). This group differed significantly from the TD group on measures of word reading and decoding skills. The DD group also demonstrated deficits in the three key phonological domains: phonological awareness (Spoonerism, phoneme segmentation, phoneme deletion), verbal short-term memory (digit span), and rapid naming (rapid automatized naming). Moreover, no differences were observed in sleep or alertness measures between the two groups.

### Procedure

**Stimuli** Non-linguistic stimuli were used, based on prior research indicating that when listeners encounter verbal material, their existing representations regarding probabilistic co-occurrences of speech sounds in their native language impact their SL performance (Siegelman et al., 2018). A further justification for using non-linguistic input was to avoid the possibility that problems in phonological processing characteristic of those with DD (Snowling, 2001) would influence their ability to extract statistical regularities from a continuous verbal input. The task and stimuli were adapted from the study conducted by Durrant et al. (2011). Stimuli consisted of sequences of pure tones taken from the Bohlen-Pierce scale (frequencies 262 Hz, 301 Hz, 345 Hz, 397 Hz, and 456 Hz) in order to avoid existing familiar tone patterns. Each tone lasted 200 ms, with a 20-ms gap between tones. Tones were sampled with a frequency of 44,100 Hz, with fixed amplitude, and were Gaussian modulated to prevent edge aliasing. Stimuli for the training/exposure session were composed of a single structured stream of 1,818 tones lasting 6 min and 40 s. Stimuli for the immediate-recall and delayed-recall sessions were composed of 168 short test

**Table 1** Series of cognitive tests and questionnaires

Ability	Test	Description
Intellectual Ability	Raven's Standard Progressive Matrices test (Raven, Court, & Raven, 1992)	Nonverbal intelligence was assessed by the Raven's-SPM test. This task requires participants to choose the item from the bottom of the figure that would complete the pattern at the top. The maximum raw score is 60. Test reliability coefficient is 0.9
Verbal Short-Term Memory	Digit Span Wechsler Adult Intelligence Scale (WAIS-III; Wechsler, 1997)	The Digit Span subtest from the Wechsler Adult Intelligence Scale. This task requires participants to recall the names of the digits presented auditorily in the order they were presented, with a maximum total raw score of 30. Task administration is discontinued after a failure to recall two trials with a similar length of digits
Naming Skills	Rapid Automatized Naming (RAN; Breznitz & Misra, 2003)	Participants are required to orally name items presented visually, as rapidly as possible. The exemplars are drawn from a constant category (RAN colors, RAN categories, RAN numerals, and RAN letters). This requires retrieval of a familiar phonological code for each stimulus and coordination of phonological and visual (color) or orthographic (letters) information quickly on time. The reliability coefficient of these tests ranges from .98 to .99
Reading Skills	One-minute test of words and One-minute test of nonwords (Shatil, 1997)	Examined by the One-Minute Test of Words (Shatil, 1995a) and the One-Minute Test of Nonwords (Shatil, E. (1995b), which assess the number of words and nonwords accurately read aloud within 1 min. The One-Minute Test of Words contains 168 non-vowelized Hebrew words of an equivalent level of difficulty, listed in columns, ranging from high to low lexical frequency. The One-Minute Test of Nonwords contains 86 successively difficult vowelized Hebrew nonwords listed in seven columns. Both accuracy (number of correct words read per minute) and speed (number of items read per minute) were measured
Phonological Processing	Phoneme segmentation test (Breznitz & Misra, 2003)	This measure assesses the participant's ability to break a word into its component phonemes. This test consists of 16 nonwords. Scores are based on the total number of accurate responses and test performance time
	Phoneme deletion test (Breznitz & Misra, 2003)	In this test, participants are required to repeat nonwords without a specific phoneme as rapidly as possible. The nonwords are presented auditorily and vary in complexity, with a maximum total raw score of 25
	Spoonerism Test (adapted from Brunswick et al., 1999)	Participants are required to switch the first syllables of two word-pairs and then synthesize the segments to provide new words. The maximum raw score is 12
Sleep Quality	Pittsburgh Sleep Quality Index (PSQI) (Buysse et al., 1989).	The PSQI questionnaire assesses sleep habits during the past month only. This questionnaire consists of nineteen individual items that generate seven "component" scores: subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disturbances, use of sleeping medication, and daytime dysfunction. In scoring the PSQI, seven component scores are derived, each scored 0 (no difficulty) to 3 (severe difficulty). The component scores are summed to produce a global score (range 0–21). Higher scores indicate poorer sleep quality
Sleepiness Self-Assessment	Stanford Sleepiness Scale (SSS) (Hoddes et al., 1973).	The Stanford Sleepiness Scale is a self-assessment questionnaire, developed by Hoddes et al. (1973). In the Stanford Sleepiness Scale participants report sleepiness at bedtime, wake time, how they feel at a given moment in a range of 1 = Feeling active, vital, alert, or wide awake to 7 = No longer fighting sleep, sleep onset soon; having dream-like thoughts, with higher scores indicating sleeper and lower scores indicating more activity

**Table 2** Demographic and psychometric data of the DD and TD groups

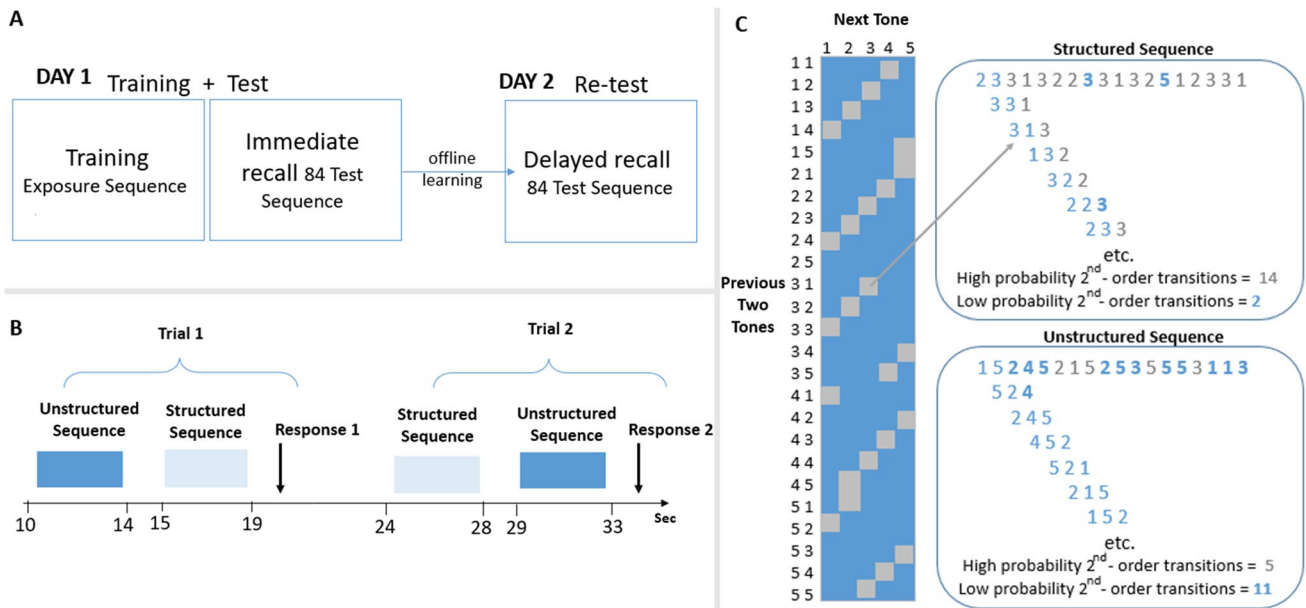
Measurement	Mean control	S.D. control	Mean dyslexia	S.D. dyslexia	t value	p	Cohen's d
Age (in years)	24.8	2.6	26.1	4.2	1.1	0.2	0.3
<b>Reading skills</b>							
Oral words recognition (accuracy)	113.6	13.3	69.2	22.0	-7.8	**	-2.4
Oral non-words recognition (accuracy)	63.6	10.8	23.1	8.5	-13.2	**	-4.1
<b>Naming skills</b>							
Naming letters (time)	22.4	2.5	25.6	3.7	3.1	**	0.9
Naming objects (time)	33.1	5.3	41.6	8.0	3.9	**	1.2
Naming numbers (time)	17.7	3.0	22.2	3.4	4.3	**	1.3
Naming colors (time)	27.7	4.6	33.9	7.1	3.2	**	1.0
<b>Phonological processing</b>							
Phoneme segmentation (time)	70.0	15.0	148.4	50.4	6.8	**	2.1
Phoneme deletion (time)	92.1	20.8	194.1	64.6	6.8	**	2.1
Spoonerism (time)	113.5	38.0	313.2	165.2	5.3	**	1.6
<b>Short verbal working memory</b>							
Digit span	12.4	2.9	10.2	3.0	-2.3	*	0.7
<b>Intellectual ability</b>							
Raven test	55.5	2.7	52.7	9.9	-1.2	.2	0.3
<b>Sleep quality</b>							
The Pittsburgh Sleep Quality Index (PSQI)	3.6	1.4	3.6	1.3	-.04	.9	0
<b>Sleep self-assessment</b>							
Stanford Sleepiness Scale (SSS) Immediate recall	2.2	0.8	2.1	0.7	-.56	0.5	0.1
Stanford Sleepiness Scale (SSS) Delayed Recall	1.9	0.6	1.9	0.6	-.01	0.9	0

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

streams, each containing 18 tones (lasting 3.96 s). Half of the test sequences had tones in a random order (unstructured condition), while the other half were determined by a transition matrix (illustrated in Fig. 1C) containing the probabilities for each potential transition between a pair of tones and the subsequent tone, forming a second-order Markov chain (structured condition).

In the transition matrix, each row-column combination has an entry that specifies the probability that the two tones associated with that row will be followed by the tone associated with that column. Each row in the transition matrix contained one high probability (termed a likely transition) ( $p = 0.9$ ; shown in grey in Fig. 1B) and four equal low probabilities (unlikely transitions) ( $p = 0.025$ ; shown in blue in Fig. 1B). This was done to make sure that any given pair of tones would be followed by a particular third tone 90% of the time, but 10% of the time would be followed by any of the other four possible tones, thus creating a probabilistic sequential structure. The transition matrix was built in such a way that equal probabilities were given for each of the five tones when considering only a single previous tone (i.e., uniform first-order transitions) or no previous tones (uniform zero-order transitions). Therefore,

any noticeable structure in the sequences was second-order or higher, requiring participants to develop sensitivity not only to a single previous tone (either one- or two-back) but rather to both previous tones. These stimuli are considered less prone to chunking compared to stimuli constructed by explicitly concatenated chunks (such as the widely used triplet paradigm) or lower order stimuli (in which recurrent pairs are more explicitly apparent). Hence, the use of such types of sequences minimizes the contamination of learning by declarative knowledge (Jiménez et al., 2006; Jiménez & Mendez, 1999; Schvaneveldt & Gomez, 1998; Song et al., 2007). Structured sequences were created by randomly sampling the transition matrix and in a way that three difficulty levels were defined (easy, medium, hard), which corresponded to different levels of structure within the sequence. This can be achieved by changing the probability of the likely transition in the transition matrix prior to sampling, with a harder difficulty level having a lower value; this manipulation reduces the number of likely transitions within a sequence, and correspondingly increases the number of unlikely transitions. Such random sampling, however, may not provide an exactly proportional number of likely and unlikely transitions within short tone sequences



**Fig. 1** (A) Timeline of the experiment: The first session was divided into two parts: a training phase and a test phase (immediate recall). After a 12-h sleep interval, participants were retested on novel but similar SL structure (delayed recall). (B) Test trials in which two pairs of short sequences were introduced. (C) Structured and unstructured sequences. *Left*: Transition matrix for the exposure stream and structured test sequences. Values are color coded probabilities, with blue = 0.025 and gray = 0.90. The row indexes the last two tones that occurred, the column indexes the next tone that could occur, and the grayscale value gives the probability of this transition. The

matrix is set up in such a way that zero- and first-order transitions are fully balanced, ensuring that they cannot provide additional structural information. *Right*: A structured sequence (top) and an unstructured sequence (bottom), showing the set of second-order transitions that make up the sequence. High-probability transitions are in gray, low-probability transitions are in blue. The structured sequence is constrained to have 14 high-probability transitions, while each transition in the unstructured sequence is generated randomly and happened to produce five high-probability transitions in this particular case

such as those used here (16-s order transitions), for example, a hard sequence might have 15 likely transitions by chance, while an easy sequence may come out with only 12. In order to make sure that all easy sequences were easier than medium sequences, all of which in turn were easier than hard sequences, the number of likely transitions were instead constrained as follows: 14 in easy sequences, 11 in medium sequences, and eight in hard sequences. This is equivalent to setting the likely transition probability to 0.875, 0.6875, and 0.5, respectively, but with any sampling error in the creation of the sequences removed.

**SL Task** The SL task was adapted from the study of Durrant et al. (2011). During the training phase participants were familiarized with the transitional probabilities. The training phase was followed by an immediate-recall session phase in which participants were required to perform a two-alternative forced-choice (2afc) task consisting of 84 trials, 28 trials for each difficulty level (Fig. 1B). Each trial consisted of a pair of two short sequences of 18 tones each: one structured, with similar probabilities as the exposure stream, and one unstructured. The participant's task was to judge which

of the two short sequences was most similar to the sounds heard during training, by pressing ‘1’ or ‘2’ on the computer keyboard to indicate the first or second sequence. The delayed-recall session consisted of a further 84 2afc trials equivalent to the immediate-recall session. The structured sequences in this session were novel but shared the transition probabilities with the exposure stream from the training session (namely they had the same statistical structure as the exposure stream). The unpredictable sequences were novel and were randomly generated.

Participants were first required to complete a background online questionnaire that included questions about gender, age, and academic background, before arriving at the laboratory. Then, participants arrived at the laboratory to complete a series of cognitive and linguistic tests in a single session before completing the SL task, which was administered in two consecutive sessions. Figure 1A presents a summary of the session timeline of the SL task administration. The first session included training and test phases (immediate recall) and the second session included a retest (delayed recall). Participants performed the first session at night and the second session in the morning, after a 12-h sleep interval. At

the beginning of each session (immediate- vs. delayed-recall sessions), participants filled out the SSS (Hoddes et al., 1973) to measure alertness.

## Statistical approach

**Power analysis** Previous research that used the SL task employed in the present study (Durrant et al., 2011; Durrant et al., 2013) revealed large effect sizes for sleep-dependent consolidation effects (i.e., averaged partial eta squared of 0.21). Furthermore, in the study of Gabay et al. (2015) a large effect size was observed when comparing DD and control participants on a similar but not identical SL task (partial eta squared of 0.25). However, because no previous study used the task employed in the current study with young adults with DD, we erred on the side of caution in predicting only medium effect sizes ( $d = 0.5$ ,  $f = 0.25$  or  $\eta_p^2 = 0.06$ ) to test within- and between-variables interactions (e.g., interactions between session and group). A power analysis (calculated using Gpower software; Faul et al., 2007) indicates that in order to detect within- and between-group interaction effects a total sample of 34 participants is needed to obtain statistical power at a 0.80 level with an alpha of 0.05. Therefore, with a total sample of 41 participants, our study was adequately powered to detect a medium effect size.

**Statistical analyses of SL** SL performance was measured by calculating the percentage of trials in which the structured sequence was correctly identified as in previous research (Durrant et al., 2011) and was used as the dependent variable for analyses reported below. First, simple t-tests were conducted to examine whether learning occurred for each group separately at above chance (50%) in the immediate-recall and delayed-recall sessions. Next, to account for differences in initial learning, performance of the two groups was compared during the immediate-recall session by using a mixed-model analysis of variance (ANOVA), with Group (DD vs. control) as between-subjects' factor and Difficulty (easy, medium, or difficult) as within-subjects' factors. To assess overnight-memory consolidation, a three-way ANOVA was conducted, with Group (DD vs. control) as between-subjects factor and Session (immediate vs. delayed recall) and Difficulty (easy, medium, or difficult) as within-subjects' factors. For the ANOVA analyses, only significant main effects or interactions are reported.

## Results

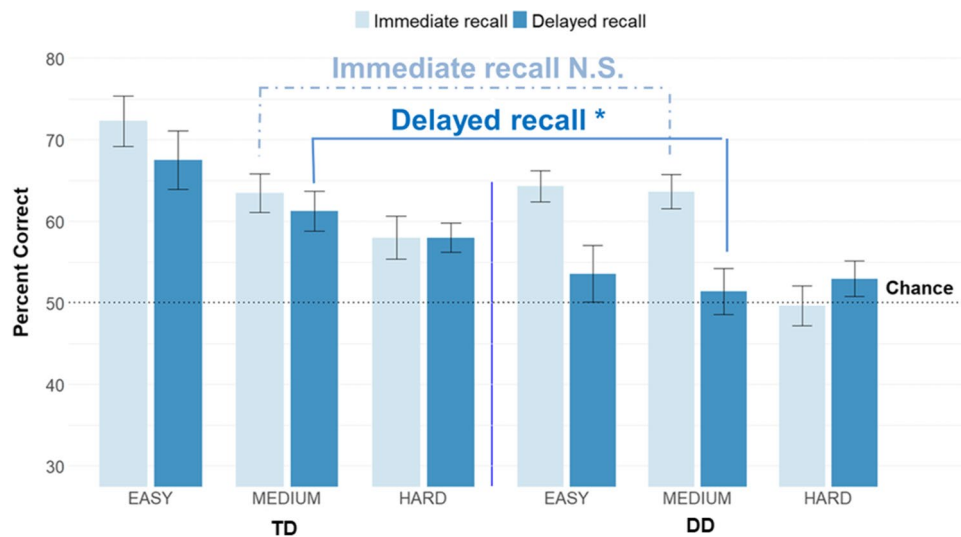
**SL performance tested against chance level** The TD group learned all SL structures above chance in the immediate-recall

session [ $t(1, 20) = 7.148$ ,  $p = .001$ ; Cohen's  $d = 1.571$  easy SL,  $t(1, 20) = 5.686$ ,  $p = .001$ ; Cohen's  $d = 1.301$  medium SL,  $t(1, 20) = 3.065$ ,  $p = .006$ ; Cohen's  $d = .636$ , hard SL] and retained that knowledge in the delayed-recall session [ $t(1, 20) = 4.871$ ,  $p = .001$ ; Cohen's  $d = 1.062$  easy SL,  $t(1, 20) = 4.593$ ,  $p = .001$ ; Cohen's  $d = 1$  medium SL,  $t(1, 20) = 4.545$ ,  $p = .001$ ; Cohen's  $d = .875$ , hard SL]. The DD group exhibited learning above chance for the easy [ $t(1, 19) = 7.504$ ,  $p = .001$ ; Cohen's  $d = 1.750$ ] and medium SL [ $t(1, 19) = 6.437$ ,  $p = .001$ ; Cohen's  $d = 1.444$ ], but not for the hard SL [ $t(1, 19) = -.154$ ,  $p = .878$ ; Cohen's  $d = -.101$ ] in the immediate-recall session, whereas none of the SL structures were retained above chance in the delayed-recall session [minimum  $p = .19$ ].

**Initial learning** The main effect of Group failed to reach significance,  $F(1, 39) = 3.857$ ,  $p = .057$ ,  $\eta_p^2 = .091$ , while a significant main effect of Difficulty was observed,  $F(2, 78) = 31.045$ ,  $p = .001$ ,  $\eta_p^2 = .443$ . The Group  $\times$  Difficulty interaction was significant,  $F(2, 78) = 3.336$ ,  $p = .041$ ,  $\eta_p^2 = .078$ . Further analysis revealed that the performance of the DD group was poorer than that of the TD group on the easy,  $F(1, 39) = 4.657$ ,  $p = .037$ ;  $\eta_p^2 = .106$  and hard SL,  $F(1, 39) = 5.486$ ,  $p = .024$ ;  $\eta_p^2 = .123$ . No group differences were observed in the medium SL,  $F(1, 39) = .004$ ,  $p = .949$ ;  $\eta_p^2 = .001$ .

**Overnight memory consolidation** The DD group in general was significantly less accurate than the TD group,  $F(1, 39) = 9.297$ ,  $p = .004$ ;  $\eta_p^2 = .192$ . There was a main effect of Session, indicating that participants were less accurate in the delayed-recalled session compared to the immediate-recall session,  $F(1, 39) = 7.799$ ,  $p = .008$ ,  $\eta_p^2 = .166$ . There was also a significant main effect of Difficulty,  $F(2, 78) = 19.666$ ,  $p = .001$ ,  $\eta_p^2 = .335$ . Further analysis suggested a significant linear trend, such that the more complex the SL structure the less accurate the listeners,  $F(1, 39) = 33.142$ ,  $p = .001$ ;  $\eta_p^2 = .489$ . The Difficulty  $\times$  Session interaction was significant,  $F(2, 78) = 9.046$ ,  $p = .001$ ,  $\eta_p^2 = .188$ . With relevance for the objective of the present study, there was a significant three-way interaction of Group  $\times$  Session  $\times$  Difficulty,  $F(2, 78) = 3.793$ ,  $p = .027$ ,  $\eta_p^2 = .088$  (see Fig. 2). To understand the basis of this interaction, we compared overnight consolidation between the DD versus TD groups using 2 (Group)  $\times$  2 (Session) ANOVAs conducted for each level separately. The Session  $\times$  Group interaction was not significant for easy SL,  $F(1, 39) = 1.705$ ,  $p = .199$ ;  $\eta_p^2 = .041$ <sup>1</sup> or hard SL conditions,  $F(1, 39) = .534$ ,  $p = .469$ ;  $\eta_p^2 = .013$ , but the DD group

<sup>1</sup> Notably, the Group  $\times$  Session interaction for the easy difficulty level was not statistically different from that of the medium difficulty level ( $F < 1$ ).



**Fig. 2** The performance of the DD and Control groups in the immediate-recall and delayed-recall sessions as a function of item difficulty. Error bars represent one standard error

showed significantly less overnight consolidation in the medium SL condition [Group  $\times$  Session:  $F(1, 39) = 7.479$ ,  $p = .009$ ;  $\eta_p^2 = .160$ .; Group differences: immediate-recall session:  $F(1, 39) = .004$ ,  $p = .949$ ;  $\eta_p^2 = .001$ ; delayed-recall session  $F(1, 39) = 6.981$ ,  $p = .011$ ;  $\eta_p^2 = .15$ ]. This interaction reflecting reduced overnight consolidation in DD participants passed Bonferroni correction and could not be attributed to a difference in initial learning since the groups did not differ at the medium difficulty level.<sup>2</sup>

## Discussion

In Experiment 1 we observed that TD learners were able to retain the acquired SL knowledge following sleep, whereas the DD group failed to do so. This finding points to the existence of an overnight consolidation-based stabilization process (task performance is maintained; Nettersheim et al., 2015) rather than to an overnight consolidation-based enhancement process (task performance is enhanced; see Walker, 2005 for a discussion of these concepts) of SL knowledge that occurred in the TD group but failed to occur in DD. Such a process can sometimes be unmasked on the behavioral level by using an interference design in which participants attempt to acquire a second task after learning an initial task (Brashers-Krug et al., 1996; Ellenbogen et al., 2006). Therefore, to directly examine the existence of a consolidation-based stabilization process in this task, we

<sup>2</sup> The pattern of results regarding memory consolidation was maintained when using percent improvement relative to initial learning (delayed recall-immediate recall/immediate recall)  $\times$  100 as the dependent variable.

conducted a second experiment in which SL was examined in TD readers using an interference design.

## Experiment 2

In Experiment 2 we tested three groups of TD readers using an interference design, which is the common way to assess consolidation-based stabilization in memory consolidation research (Brashers-Krug et al., 1996).

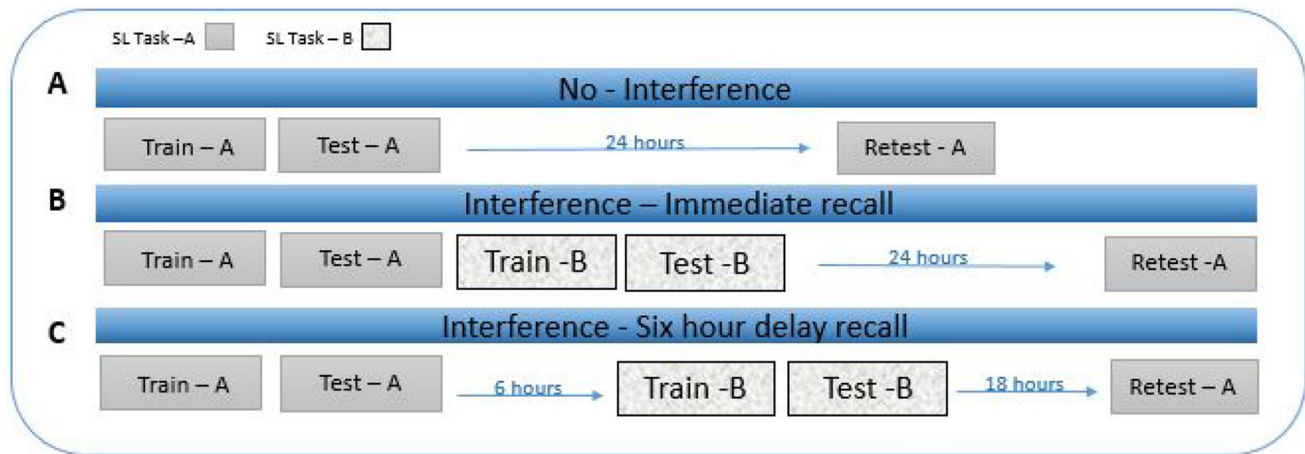
## Methods

**Participants** Young neurotypical adult participants (33 total; six males and 27 females,  $M_{age} = 25.18$  years,  $SD = 2.65$  years) were recruited in person and assigned randomly to one of three conditions: (1) *no interference* ( $N = 10$ ); (2) *immediate interference* ( $N = 13$ ); (3) *6-h delay interference* ( $N = 10$ ). The study was conducted at the University of Haifa in accordance with the Declaration of Helsinki and participants were compensated for their participation.

**Stimuli** Stimuli in the first SL task were similar to those of Experiment 1. A different set of stimuli was used for the second SL task.

**Procedure** Three groups of participants performed the same SL task described in Experiment 1 during a first testing session (immediate recall) and all were retested after a 24-h interval that included sleep (delayed recall) (see Fig. 3). The first group performed only the original SL task - A in the first session (no interference). The second group performed the SL





**Fig. 3** Study protocol. Participants were randomly assigned to one of three groups: (A) No interference: participants performed only the SL task-A in the first session without interference. (B) Immediate interference: Participants performed the SL task-A and then, immediately

after, performed a different SL task-B, in the first session. (C) Six-hour delay interference: participants performed the SL task-A and after a 6-h interval performed the second SL task-B. All groups were retested on the initial SL task-A after a 24-h interval

task - A and then, immediately after, a second SL task - B, still in the first session (immediate interference). The third group performed the SL task (A) and after a 6-h interval performed the second SL task (B) (6-h interval interference). All groups were tested in the second session on sequences with the same statistical structure as in the first task (SL task - A).

### Statistical approach

**Power analysis** Previous research that used the SL task employed in the present study (Durrant et al., 2011; Durrant et al., 2013) revealed large effect sizes for sleep-dependent consolidation effects (i.e., averaged partial eta squared of 0.21). However, since consolidation of auditory SL was not previously studied using an interference design, we again erred on the side of caution and predicted medium effect sizes ( $d = 0.5$ ,  $f = 0.25$  or  $\eta_p^2 = 0.06$ ) to test within- and between-variables interactions (e.g., interactions between session and group). A power analysis (calculated using Gpower software; Faul et al., 2007) indicates that in order to detect within- and between-group interaction effects, a total sample of 42 participants is needed to obtain statistical power at a 0.80 level with an alpha of 0.05. This suggests that the study was slightly underpowered to detect medium (but not large) effect sizes.

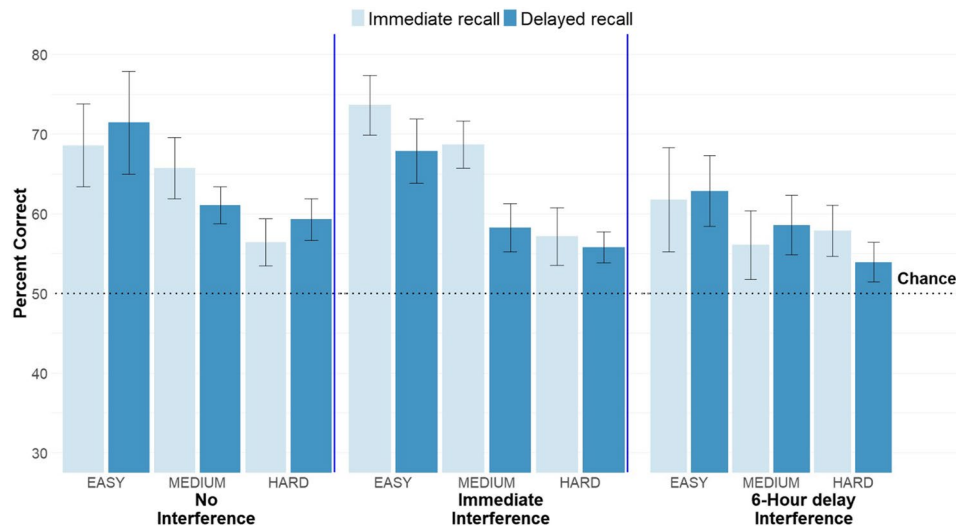
**Statistical analyses of SL** SL performance was measured by calculating the percent number of trials in which the structured sequence was correctly identified (Durrant et al., 2011). To assess group differences in memory consolidation a mixed ANOVA was conducted, with Group (no interference, immediate interference, 6-h delay interference) as the between-subjects factor and Session (immediate vs. delayed recall) and

Difficulty (easy, medium, and difficult) as within-subject factors. In what follows only significant main effects or interactions are reported.

### Results and discussion

A significant main effect of Difficulty was found,  $F(2, 60) = 16.01$ ,  $p = .001$ ,  $\eta_p^2 = .347$ . Further analysis suggested a significant linear trend, such that the more complex the SL structure, the less accurate the listeners,  $F(1, 30) = 24.419$ ,  $p = .001$ ;  $\eta_p^2 = .282$ . The Group  $\times$  Session interaction was significant,  $F(2, 30) = 3.372$ ,  $p = .048$ ,  $\eta_p^2 = .183$  (see Fig. 4). Further analysis suggested that while the baseline (no interference) group and the 6-h interference group were capable of retaining their performance in the delayed-recall session compared to the immediate-recall session (all  $F_s < 1$ ), the immediate interference group showed decreased performance in the delayed compared to the immediate-recall session,  $F(1, 30) = 10.645$ ,  $p = .002$ ;  $\eta_p^2 = .06^3$ . Furthermore, there was no

<sup>3</sup> Notably this finding is less likely to be attributed to initial learning differences across the three groups. When initial learning was tested by using a mixed-ANOVA, with Group (DD vs. control) as the between-subjects factor and Difficulty (easy, medium, or difficult) as the within-subjects factor on percent accuracy during immediate recall, only a significant main effect of Difficulty was observed ( $F(2, 60) = 7.80$ ,  $p = .01$ ;  $\eta_p^2 = .20$ ). No significant main effect of Group was observed,  $F(2, 30) = 1.73$ ,  $p = .19$ ;  $\eta_p^2 = .103$ . Three one-way ANOVAs conducted separately for each difficulty level did not reveal any significant differences in the easy, medium, or difficult SL during the immediate-recall session between the groups. Therefore, no differences in initial learning were observed across the groups. This reduces the possibility that initial learning differences could account for the observed effect.



**Fig. 4** The performance of the three different groups in the immediate-recall and delayed-recall sessions as a function of session and item difficulty. Error bars represent one standard error

difference in the amount of retention between the control (no interference group) and the 6-h delay group,  $F(1, 30) = .027$ ,  $p = .871$ ;  $\eta_p^2 = .001$ , that is, after 6 h had passed the SL in task A was not disrupted when task B was learned. Overall, these findings suggest that the group that performed a different SL task immediately after performing the initial SL task failed to maintain the acquired knowledge of the initial SL when tested after a sleep interval. Therefore, immediately after training, memory of SL knowledge is highly susceptible to interference but becomes resilient to disruption after a 6-h delay when retested after sleep and is equivalent to that observed when no interference is induced. The existence of consolidation effects of SL knowledge can therefore be unmasked by using an interference design, consistent with prior research (Brashers-Krug et al., 1996; Nettersheim et al., 2015). These findings suggest the need for an offline temporal time window for the stabilization of SL knowledge. Power analysis indicated that the study (Experiment 2) was slightly underpowered to detect a medium (but not large) effect size and that had null findings been obtained this may have been due to insufficient power. However, we observed significant group differences between the immediate interference group as compared to the no interference group and the 6-delay interference group, consistent with our hypothesis and with the previous large effect sizes seen with this task, which our sample was sufficient to detect. In the case that our study was not sensitive enough to detect differences between the 6-delay interference group versus no interference group, this does not change our conclusion of the existence of consolidation-based stabilization effects (differences observed between the immediate interference group vs. other groups). Nevertheless, these findings should be further replicated and tested in larger studies.

## General discussion

The current results show that SL knowledge originating from passive exposure was consolidated into long-term memory in TD readers after the initial exposure was concluded. In particular, the TD group could maintain the learned SL information, as indicated by similar SL performance in both the immediate-recall and delayed-recall sessions that exceeded chance level. This pattern of results can point to the involvement of a consolidation-based stabilization stage in which no additional gains in performance are seen after initial acquisition but in which the behavioral performance is maintained after an offline interval (Nettersheim et al., 2015; Walker, 2005). This is further supported by the results of Experiment 2, in which TD readers could maintain the same level of SL performance only if there was a 6-h interval between the learned SL and the learning of new SL information. Such findings corroborate previous research pointing to a wake-based stabilization phase in motor skill learning using an interference design (Brashers-Krug et al., 1996).

In contrast to the TD readers group, the DD group was not capable of maintaining the acquired SL knowledge after a sleep interval. In particular, whereas no group differences were observed for medium SL structures in the immediate-recall session, a significant group difference was evident in the delayed-recall session. This can be attributed to a failure in the overnight consolidation-based stabilization process in DD. One thing to consider, however, is why significant group differences in SL performance across sessions were not evident at the easy and hard difficulty levels. First, during the immediate-recall

session, the SL performance of the DD group at the hard difficulty level did not exceed chance, so there was no initial learning from which consolidation processes could take place. Second, although significant group differences in SL performance were not observed across sessions at the easy difficulty level, the pattern of results was very similar and was not statistically different than that observed at the medium difficulty level ( $F < 1$ ).

It is important to note that the DD group performed significantly worse than the TD group during the immediate-recall session of the easy and hard difficulty SL levels, indicating problems in initial learning of SL knowledge. This is consistent with previous findings revealing segmentation difficulties in DD during initial learning (Gabay et al., 2015; Sigurdardottir et al., 2017). The pattern of results raises the possibility that problems in initial learning rather than a memory consolidation deficit could influence the ability of people with DD to retain SL knowledge across time. The observation that both groups performed similarly at the medium difficulty SL level should be interpreted with caution, as the SL performance of the DD group was impaired relative to the TD group on the other difficulty levels. Yet the possibility that initial learning resulted in less retention of SL knowledge in the DD group is less probable, since additional analysis indicated that poorer initial learners across both groups exhibited greater consolidation-based stabilization (see Online Supplementary Materials). These findings, in combination with a lack of group differences in SL performance at the medium difficulty level, imply the involvement of a memory consolidation deficit rather than the mere influence of an initial learning deficit on the retention abilities of people with DD. Future studies will be necessary to identify the contribution of initial learning impairments vs. memory consolidation to the retention deficits observed in individuals with DD.

Since our protocol included a sleep interval, this may point to the possibility of an impairment in sleep-dependent consolidation processes in the DD group. However, future studies are needed to determine whether the SL deficit observed in the current study is specific to sleep, by using a wake versus sleep design or measuring sleep parameters with polysomnography. The present findings resonate with prior research pointing to sleep-dependent memory consolidation deficits of linguistic information in DD (Reda et al., 2021; Smith et al., 2018), and broaden these findings to the non-verbal auditory perceptual domain. Our findings are therefore difficult to reconcile with theories positing that DD arises due to deficits in phonological processing (Snowling, 2001). The observed deficiency in DD across an acoustic domain implies an impairment affecting domain-general processes, though future studies are needed to determine

whether the observed pattern arises from a domain-general or a domain-specific deficiency (Singh & Conway, 2021).

In this regard, the paradigm we used is especially relevant with regard to a major theoretical framework in the field of DD (Nicolson & Fawcett, 2011; Ullman et al., 2020). According to the Procedural Deficit Hypothesis, a dysfunction to the striatal memory system could account for the linguistic symptoms of people with DD. Intriguingly the present paradigm has been shown to engage the procedural memory system during a consolidation phase among neurotypicals (Durrant et al., 2013). Furthermore, although the paradigm involves explicit familiarity judgments, the probabilistic structure of the task reduced the ability to chunk information, which is related to declarative processing. Therefore, based on both the brain-based and computation-based approaches (Bogaerts et al., 2020), it may be argued that the consolidation SL deficit observed in the present study may be attributed to a procedural memory dysfunction in DD.

Our results suggest that it is important to consider different types of SL knowledge in those with DD and that the impairments observed in DD cannot be conceived as all or none (Arciuli & Conway, 2018). Rather, it seems that several types of SL information could still be learned by people with DD during the encoding phase (though to a lesser degree than TD readers), while learning more complex SL structures presents a greater source of challenge for those with DD (Lum et al., 2013). Our study also highlights the importance of studying different stages of learning in those with DD. Although people with DD might be able to learn some forms of SL as TD readers during an acquisition phase, they may fail to consolidate that knowledge into long-term memory. Therefore, previous studies examining SL in DD using the segmentation task in one training session (van Witteloostuijn et al., 2019) might have underestimated the SL deficit we observed in the current study. Although prior research examined SL in DD across time (Gabay et al., 2012a; Hedenius et al., 2013; Hedenius et al., 2020), our study is the first to report a failure to consolidate segmentation-based SL knowledge in those with DD. Since language learning critically depends on such domain-general learning capacities (Saffran & Thiessen, 2007), a deficit in retrieving segmentation-based SL knowledge may have a negative impact on the ability of people with DD to form robust linguistic representations. Such an impairment is likely to place learners with DD in constant need of relearning, thus influencing their ability to retain segmented units from fluent speech in order to retrieve statistically structured phonological information, including the ability to retain statistical regularities embedded in sound categories over time.

In sum, in the present study TD and DD participants were exposed to auditorily statistically structured information and

were retested after a sleep interval. Despite evidence for consolidation-based stabilization of auditory SL knowledge in the TD group, the DD group failed to consolidate auditory SL knowledge over a sleep interval. Such a failure is likely to have a negative impact on the ability of people with DD to form precise linguistic representations in long-term memory.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.3758/s13423-022-02169-y>.

**Acknowledgements** This research was supported by a grant from the Israel Science Foundation (grant No. 734/22) awarded to Yafit Gabay, by the Binational Scientific Foundation (grant No. 2015227) and the National Science Foundation-Binational Scientific Foundation (BSF2016867, NSF BCS1655126) grants awarded to Yafit Gabay and Lori Holt, respectively and by the Simons Foundation Autism Research Initiative Grant (875476) awarded to Dara Manoach. This study is part of the research conducted at the University of Haifa by Ranin Ballan, as partial fulfillment of her requirements for a PhD degree under the supervision of Yafit Gabay.

## References

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.).
- Arciuli, J., & Conway, C. M. (2018). The promise—And challenge—Of statistical learning for elucidating atypical language development. *Current Directions in Psychological Science*, 27(6), 492–500.
- Arciuli, J., & Simpson, I. C. (2012). Statistical learning is related to reading ability in children and adults. *Cognitive Science*, 36(2), 286–304.
- Baran, B., Correll, D., Vuper, T. C., Morgan, A., Durrant, S. J., Manoach, D. S., & Stickgold, R. (2018). Spared and impaired sleep-dependent memory consolidation in schizophrenia. *Schizophrenia Research*, 199, 83–89.
- Ben-Zion, D., Gabitov, E., Prior, A., & Bitan, T. (2022). Effects of sleep on language and motor consolidation: Evidence of domain general and specific mechanisms. *Neurobiology of Language*, 3(2), 180–213. [https://doi.org/10.1162/nol\\_a\\_00060](https://doi.org/10.1162/nol_a_00060)
- Bogaerts, L., Siegelman, N., & Frost, R. (2020). *Statistical learning and language impairments: Toward more precise theoretical accounts* (p. 1745691620953082). Perspectives on Psychological Science.
- Bogaerts, L., Siegelman, N., & Frost, R. (2021). Statistical learning and language impairments: Toward more precise theoretical accounts. *Perspectives on Psychological Science*, 16(2), 319–337.
- Bogaerts, L., Szmalec, A., Hachmann, W. M., Page, M. P., & Duyck, W. (2015). Linking memory and language: Evidence for a serial-order learning impairment in dyslexia. *Research in Developmental Disabilities*, 43, 106–122.
- Brashers-Krug, T., Shadmehr, R., & Bizzi, E. (1996). Consolidation in human motor memory. *Nature*, 382(6588), 252–255.
- Breznitz, Z., & Misra, M. (2003). Speed of processing of the visual-orthographic and auditory-phonological systems in adult dyslexics: The contribution of “asynchrony” to word recognition deficits. *Brain and Language*, 85(3), 486–502.
- Brunswick, N., McCrory, E., Price, C. J., Frith, C. D., & Frith, U. (1999). Explicit and implicit processing of words and pseudowords by adult developmental dyslexics. *Brain*, 122(10), 1901–1917. <https://doi.org/10.1093/brain/122.10.1901>
- Buysse, D. J., Reynolds III, C. F., Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research. *Psychiatry Research*, 28(2), 193–213.
- Censor, N., & Sagi, D. (2008). Benefits of efficient consolidation: Short training enables long-term resistance to perceptual adaptation induced by intensive testing. *Vision Research*, 48(7), 970–977.
- Conway, C. M. (2020). How does the brain learn environmental structure? Ten core principles for understanding the neuro-cognitive mechanisms of statistical learning. *Neuroscience & Biobehavioral Reviews*, 112, 279–299.
- Davis, M. H., Di Betta, A. M., Macdonald, M. J., & Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *Journal of Cognitive Neuroscience*, 21(4), 803–820.
- Démonet, J.-F., Taylor, M. J., & Chaix, Y. (2004). Developmental dyslexia. *The Lancet*, 363(9419), 1451–1460.
- Dudai, Y., Karni, A., & Born, J. (2015). The consolidation and transformation of memory. *Neuron*, 88(1), 20–32.
- Durrant, S. J., Cairney, S. A., & Lewis, P. A. (2013). Overnight consolidation aids the transfer of statistical knowledge from the medial temporal lobe to the striatum. *Cerebral Cortex*, 23(10), 2467–2478.
- Durrant, S. J., Cairney, S. A., & Lewis, P. A. (2016). Cross-modal transfer of statistical information benefits from sleep. *cortex*, 78, 85–99.
- Durrant, S. J., Taylor, C., Cairney, S., & Lewis, P. A. (2011). Sleep-dependent consolidation of statistical learning. *Neuropsychologia*, 49(5), 1322–1331.
- Earle, F. S., & Myers, E. B. (2015). Sleep and native language interference affect non-native speech sound learning. *Journal of Experimental Psychology: Human Perception and Performance*, 41(6), 1680.
- Earle, S., & Myers, E. (2013). The effect of sleep on learned sensitivity to a non-native phonetic contrast. *The Journal of the Acoustical Society of America*, 134(5), 4107–4107.
- Ellenbogen, J. M., Hulbert, J. C., Stickgold, R., Dinges, D. F., & Thompson-Schill, S. L. (2006). Interfering with theories of sleep and memory: Sleep, declarative memory, and associative interference. *Current Biology*, 16(13), 1290–1294.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G\* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191.
- Fawcett, A. J., & Nicolson, R. I. (2019). Development of dyslexia: The delayed neural commitment framework. *Frontiers in Behavioral Neuroscience*, 13, 112.
- Finn, A. S., & Hudson Kam, C. L. (2015). Why segmentation matters: Experience-driven segmentation errors impair “morpheme” learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(5), 1560.
- Gabay, Y. (2021). Delaying feedback compensates for impaired reinforcement learning in developmental dyslexia. *Neurobiology of Learning and Memory*, 185, 107518.
- Gabay, Y., Schiff, R., & Vakil, E. (2012a). Attentional requirements during acquisition and consolidation of a skill in normal readers and developmental dyslexics. *Neuropsychology*, 26(6), 744.
- Gabay, Y., Schiff, R., & Vakil, E. (2012b). Dissociation between the procedural learning of letter names and motor sequences in developmental dyslexia. *Neuropsychologia*, 50(10), 2435–2441.
- Gabay, Y., Thiessen, E. D., & Holt, L. L. (2015). Impaired statistical learning in developmental dyslexia. *Journal of Speech, Language, and Hearing Research*, 58(3), 934–945.
- Giustolisi, B., & Emmorey, K. (2018). Visual statistical learning with stimuli presented sequentially across space and time in deaf and hearing adults. *Cognitive Science*, 42(8), 3177–3190.
- Hedenius, M., Lum, J. A., & Bölte, S. (2020). Alterations of procedural memory consolidation in children with developmental dyslexia. *Neuropsychology*.
- Hedenius, M., Lum, J. A., & Bölte, S. (2021). Alterations of procedural memory consolidation in children with developmental dyslexia. *Neuropsychology*, 35(2), 185.

- Hedenius, M., Persson, J., Alm, P. A., Ullman, M. T., Howard Jr., J. H., Howard, D. V., & Jennische, M. (2013). Impaired implicit sequence learning in children with developmental dyslexia. *Research in Developmental Disabilities, 34*(11), 3924–3935.
- Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., & Dement, W. C. (1973). Quantification of sleepiness: A new approach. *Psychophysiology, 10*(4), 431–436.
- Hollander, C., & Adi-Japha, E. (2021). Second graders' Grapho-motor skill learning and intentional verbal learning through repetitions: The effects of socio-educational background factors. *Frontiers in Psychology, 4250*.
- Horváth, K., Török, C., Pesthy, O., Nemeth, D., & Janacsek, K. (2020). Divided attention does not affect the acquisition and consolidation of transitional probabilities. *Scientific Reports, 10*(1), 1–14.
- Howard Jr., J. H., Howard, D. V., Japikse, K. C., & Eden, G. F. (2006). Dyslexics are impaired on implicit higher-order sequence learning, but not on implicit spatial context learning. *Neuropsychologia, 44*(7), 1131–1144.
- Inácio, F., Faísca, L., Forkstam, C., Araújo, S., Bramão, I., Reis, A., & Petersson, K. M. (2018). Implicit sequence learning is preserved in dyslexic children. *Annals of Dyslexia, 68*(1), 1–14.
- Janacsek, K., & Nemeth, D. (2012). Predicting the future: From implicit learning to consolidation. *International Journal of Psychophysiology, 83*(2), 213–221.
- Jiménez, L., & Mendez, C. (1999). Which attention is needed for implicit sequence learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(1), 236.
- Jiménez, L., Vaquero, J. M., & Lupiáñez, J. (2006). Qualitative differences between implicit and explicit sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(3), 475.
- Kahta, S., & Schiff, R. (2019). Deficits in statistical learning of auditory sequences among adults with dyslexia. *Dyslexia, 25*(2), 142–157.
- Karuz, E. A., Newport, E. L., Aslin, R. N., Starling, S. J., Tivarus, M. E., & Bavelier, D. (2013). The neural correlates of statistical learning in a word segmentation task: An fMRI study. *Brain and Language, 127*(1), 46–54.
- Kelly, S. W., Griffiths, S., & Frith, U. (2002). Evidence for implicit sequence learning in dyslexia. *Dyslexia, 8*(1), 43–52.
- Kim, R., Seitz, A., Feenstra, H., & Shams, L. (2009). Testing assumptions of statistical learning: Is it long-term and implicit? *Neuroscience Letters, 461*(2), 145–149.
- Lee, S. M.-K., Cui, Y., & Tong, S. X. (2022). Toward a model of statistical learning and reading: Evidence from a meta-analysis. *Review of Educational Research, 00346543211073188*.
- Lum, J. A., Ullman, M. T., & Conti-Ramsden, G. (2013). Procedural learning is impaired in dyslexia: Evidence from a meta-analysis of serial reaction time studies. *Research in Developmental Disabilities, 34*(10), 3460–3476.
- Nettersheim, A., Hallschmid, M., Born, J., & Diekelmann, S. (2015). The role of sleep in motor sequence consolidation: Stabilization rather than enhancement. *Journal of Neuroscience, 35*(17), 6696–6702.
- Nicolson, R. I., Fawcett, A., Brookes, R., & Needle, J. (2010). Procedural learning and dyslexia. *Dyslexia, 16*(3), 194–212.
- Nicolson, R. I., & Fawcett, A. J. (2011). Dyslexia, dysgraphia, procedural learning and the cerebellum. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*.
- Qi, Z., Sanchez Araujo, Y., Georgan, W. C., Gabrieli, J. D., & Arciuli, J. (2018). Hearing matters more than seeing: A cross-modality study of statistical learning and reading ability. *Scientific Studies of Reading, 1*–15.
- Raven, J. C., & Court, J. H. (1992). *Raven's progressive matrices and vocabulary scales*. Oxford psychologists Press.
- Raven, J. C., Court, J. H., & Raven, J. (1992). *Standard progressive matrices*. Oxford University Press.
- Reber, P. J., Gitelman, D. R., Parrish, T. B., & Mesulam, M. M. (2003). Dissociating explicit and implicit category knowledge with fMRI. *Journal of Cognitive Neuroscience, 15*(4), 574–583.
- Reda, F., Gorgoni, M., D'Atri, A., Scarpelli, S., Carpi, M., Di Cola, E., & De Gennaro, L. (2021). Sleep-related declarative memory consolidation in children and adolescents with developmental dyslexia. *Brain Sciences, 11*(1), 73.
- Robertson, E. M., Pascual-Leone, A., & Miall, R. C. (2004). Current concepts in procedural consolidation. *Nature Reviews Neuroscience, 5*(7), 576–582.
- Rüsseler, J., Gerth, I., & Münte, T. F. (2006). Implicit learning is intact in adult developmental dyslexic readers: Evidence from the serial reaction time task and artificial grammar learning. *Journal of Clinical and Experimental Neuropsychology, 28*(5), 808–827.
- Saffran, J. R. (2018). Statistical learning as a window into developmental disabilities. *Journal of Neurodevelopmental Disorders, 10*(1), 1–5.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926–1928.
- Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science, 8*(2), 101–105.
- Saffran, J. R., & Thiessen, E. D. (2007). *Domain-general learning capacities*.
- Saltzman, D., & Myers, E. (2021). Listeners are initially flexible in updating phonetic beliefs over time. *Psychonomic Bulletin & Review, 1*–11.
- Sawi, O. M., & Rueckl, J. (2019). Reading and the neurocognitive bases of statistical learning. *Scientific Studies of Reading, 23*(1), 8–23.
- Schmalz, X., Altoè, G., & Mulatti, C. (2017). Statistical learning and dyslexia: A systematic review. *Annals of Dyslexia, 67*(2), 147–162.
- Schvaneveldt, R. W., & Gomez, R. L. (1998). Attention and probabilistic sequence learning. *Psychological Research, 61*(3), 175–190.
- Shatil, E. (1995a). *One-minute test for pseudowords*. Unpublished test. University of Haifa.
- Shatil, E. (1995b). *One-minute test for regular words*. Unpublished test. University of Haifa.
- Shatil, E. (1997). *Predicting reading ability: Evidence for cognitive modularity*. Unpublished doctoral dissertation, University of Haifa.
- Siegelman, N., Bogaerts, L., Elazar, A., Arciuli, J., & Frost, R. (2018). Linguistic entrenchment: Prior knowledge impacts statistical learning performance. *Cognition, 177*, 198–213.
- Sigurdardottir, H. M., Danielsdottir, H. B., Gudmundsdottir, M., Hjartarson, K. H., Thorarinsdottir, E. A., & Kristjánsson, Á. (2017). Problems with visual statistical learning in developmental dyslexia. *Scientific Reports, 7*(1), 606.
- Singh, S., & Conway, C. M. (2021). *Unraveling the interconnections between statistical learning and dyslexia: A review of recent empirical studies*. *Frontiers in human neuroscience* (p. 15).
- Singh, S., Walk, A. M., & Conway, C. M. (2018). Atypical predictive processing during visual statistical learning in children with developmental dyslexia: An event-related potential study. *Annals of Dyslexia, 68*(2), 165–179.
- Smith, F. R., Gaskell, M. G., Weighall, A. R., Warmington, M., Reid, A. M., & Henderson, L. M. (2018). Consolidation of vocabulary is associated with sleep in typically developing children, but not in children with dyslexia. *Developmental Science, 21*(5), e12639.
- Snowling, M. J. (2001). From language to reading and dyslexia. *Dyslexia, 7*(1), 37–46.
- Song, S., Howard, J. H., & Howard, D. V. (2007). Implicit probabilistic sequence learning is independent of explicit awareness. *Learning & Memory, 14*(3), 167–176.
- Spencer, M., Kaschak, M. P., Jones, J. L., & Lonigan, C. J. (2015). Statistical learning is related to early literacy-related skills. *Reading and Writing, 28*(4), 467–490.

- Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature*, 437(7063), 1272–1278.
- Stickgold, R., James, L., & Hobson, J. A. (2000). Visual discrimination learning requires sleep after training. *Nature Neuroscience*, 3(12), 1237–1238.
- Stoodley, C. J., Harrison, E. P., & Stein, J. F. (2006). Implicit motor learning deficits in dyslexic adults. *Neuropsychologia*, 44(5), 795–798.
- Tong, S., Zhang, P., & He, X. (2020). Statistical learning of orthographic regularities in Chinese children with and without dyslexia. *Child Development*, 91(6), 1953–1969.
- Tong, X., Leung, W. W. S., & Tong, X. (2019). Visual statistical learning and orthographic awareness in Chinese children with and without developmental dyslexia. *Research in Developmental Disabilities*, 92, 103443.
- Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, 97(2), B25–B34.
- Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience*, 21(10), 1934–1945.
- Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, 92(1-2), 231–270.
- Ullman, M. T., Earle, F. S., Walenski, M., & Janacek, K. (2020). The neurocognition of developmental disorders of language. *Annual Review of Psychology*, 71, 389–417.
- van Witteloostuijn, M., Boersma, P., Wijnen, F., & Rispens, J. (2019). Statistical learning abilities of children with dyslexia across three experimental paradigms. *PLoS One*, 14(8), e0220041.
- Vicari, S., Finzi, A., Menghini, D., Marotta, L., Baldi, S., & Petrosini, L. (2005). Do children with developmental dyslexia have an implicit learning deficit? *Journal of Neurology, Neurosurgery & Psychiatry*, 76(10), 1392–1397.
- Vihman, M. (2009). Word learning and the origins of phonological systems. In *Language acquisition* (pp. 15–39). Springer.
- Walker, M. P. (2005). A refined model of sleep and the time course of memory formation. *Behavioral and Brain Sciences*, 28(1), 51–64.
- Wechsler, D. (1997). *Wechsler adult intelligence scale-(WAIS-3) San Antonio*. TX: Harcourt Assessment.
- Weiss, Y., Katzir, T., & Bitan, T. (2015). The effects of orthographic transparency and familiarity on reading Hebrew words in adults with and without dyslexia. *Annals of Dyslexia*, 65(2), 84–102.

**Open practice statement** The data and materials will be provided on request. The experiments were not preregistered.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.