**Intact Forms of Procedural Memory in Adults with Dyslexia**

Ola Ozernov-Palchik, Zhenghan Qi, Sara D. Beach, John D.E. Gabrieli

**Abstract**

Developmental dyslexia is characterized by reading difficulty and is known to be associated with atypical brain functions. One neuropsychological theory of dyslexia posits that dyslexia reflects a deficit in procedural memory (learning that is independent of the brain structures that support declarative or explicit memory). Here we examined multiple forms of procedural-learning related tasks in adults with developmental dyslexia. Adults with dyslexia exhibited typical learning on two perceptual-motor tasks that have been well-established as reflecting purely procedural memory and dependent on basal ganglia and cerebellar structures, mirror tracing and rotary pursuit. They also exhibited typical statistical learning for visual material, but impaired statistical learning for auditory material. Auditory statistical learning proficiency correlated positively with single-word reading (decoding) performance across all participants and within the group with dyslexia, linking the major difficulty in dyslexia with impaired auditory statistical learning. These findings dissociate multiple forms of procedural memory that are intact in dyslexia from a specific impairment in auditory statistical learning that is associated with reading difficulty.

Key Words: dyslexia, procedural memory, statistical learning, mirror tracing, rotary pursuit

**Introduction**

Developmental dyslexia is characterized by inaccurate and dysfluent reading. Although individuals with dyslexia most consistently demonstrate difficulties with storing, retrieving, and/or manipulating phonological representations (Brady et al., 1983; Shankweiler et al., 1979; Snowling, 2000; Stanovich & Siegel, 1994; Vellutino et al., 1994), a range of non-linguistic deficits in the motor and perceptual domains have also been reported for this population (for reviews see Folia et al., 2008; West et al., 2020; Lum et al., 2013; Nicolson & Fawcett, 2011). Multiple theories have been developed to explain the neurocognitive underpinnings of dyslexia. The *procedural deficit hypothesis* proposes that dyslexia is caused by deficits in the *procedural memory system* (or knowing how), which is critically involved in extracting and learning the sequential or distributional regularities in language (e.g., Nicolson & Fawcett, 2007, 2011; Ullman, 2004; Ullman & Pierpont, 2005). In contrast to the declarative memory system that refers to conscious memory for facts and events (or knowing that), the *procedural memory* system refers to the implicit (i.e., unconscious) learning of skills, rules, or patterned regularities in stimuli (Cohen & Squire, 1980). Procedural learning is often used interchangeably with other terms such as *implicit learning* and *statistical learning* (Berry & Dienes, 1993; Perruchet & Pacton, 2006; Shanks, 2005; Christiansen, 2019). The procedural deficit hypothesis is supported by the high rates of comorbidity of dyslexia with other disorders (Ramus et al., 2003; Wimmer et al., 1999; Boada et al., 2012), suggesting a broad deficit in learning that extends beyond the language system. In the current study, we tested whether there are domain-general procedural deficits in developmental dyslexia in adults.

The distinction between procedural memory and declarative memory in human learning arose from studies of intact skill learning in amnesic patients with severe impairments in declarative memory (Cohen and Squire, 1980). These patients had injuries to medial temporal-lobe or diencephalic brain regions. For example, the amnesic patient H.M. demonstrated intact learning skills for mirror tracing across days (Milner, 1962) and a year (Gabrieli, Corkin, Mickel, & Growdon, 1993) and for rotary pursuit (Corkin, 1968), but impaired declarative memory for the episodes in which he had learned the skills. Similar intact skill learning was shown by memory-impaired patients with Alzheimer’s disease on the same two tasks (Gabrieli, Corkin, Mickel, & Growdon, 1993; Heindel, Salmon, Shults, Walicke, & Butters, 1989), supporting the dissociation of the procedural memory system from the declarative system. Research with other neurological patient groups has suggested that such forms of procedural memory are dependent on the basal ganglia for rotary pursuit (Heindel et al., 1989; Gabrieli et al., 1997) and the cerebellum for mirror tracing (LaForce & Doyon, 2001). These studies also indicate that procedural memory is not a unitary neurobiological construct, but rather that different forms of procedural memory depend on different neural substrates.

Mirror tracing learning has been examined only once in children with dyslexia, who were slower in performance but demonstrated typical learning across trials (Vicari et al., 2005). Rotary pursuit has not been examined in dyslexia. Two other forms of learning, however, have been studied more extensively in dyslexia: serial reaction time (SRT) and statistical learning (SL).

SRT is a spatio-motor skill learning task in which participants typically see four horizontal spatial locations on a monitor and are instructed to press the corresponding button (from among four horizontal buttons) as quickly as possible. In some blocks, the order of stimulus locations follows a sequential pattern. In other blocks, stimulus locations are presented randomly. Amnesic patients show normal learning of the repeating stimulus sequence as evidenced by faster reaction times (Nissen and Bullemer, 1987) and typical participants can exhibit skill learning for the repeated sequence without declarative memory for the sequence (although those typical participants who do develop declarative memory for the sequence show greater learning) (Willingham et al., 1989). SRT experiment findings, however, vary substantially by the nature of the repeating stimulus sequence because such sequences vary in their attentional demands and susceptibility for the development of declarative memory for the sequences, both of which can influence learning (e.g., Willingham et al., 1989; Cohen et al.,1990).

There are multiple reports of both intact and impaired SRT learning in dyslexia; a meta-analysis indicates that there appears to be a deficit in SRT learning in dyslexia (Lum et al., 2013; West et al., 2021). It is difficult to synthesize these findings because of the evidence that variation in SRT paradigms in relation attentional demands and susceptibility to the influence of declarative memory may invoke cognitive and declarative memory processes beyond procedural memory. Further, there is some evidence that the degree of SRT deficit appears to be more prominent in a task involving letters, compared to a task involving nonlinguistic visual stimuli, suggesting spatio-motor sequence learning in dyslexic individuals might be constrained by separate underlying learning systems across linguistic vs. non-linguistic domains (Gabay, Schiff, & Vakil, 2012).

*Statistical learning* *(SL)*, is a form of implicit learning that plays a fundamental role in the perception and categorization of environmental inputs. Learners are thought to automatically extract the co-occurring patterns of exemplars embedded in the sensory inputs mostly through passive exposure. The modern theoretical frameworks often introduce SL as a widely defined construct that encompasses a range of incidental learning paradigms (Thiessen, 2017; Frost et al., 2019; Conway, 2020; Bogaerts et al., 2020), including category learning, SRT, artificial grammar learning, and embedded pattern learning. However, the relative contributions of declarative and procedural memory to most SL task performances are largely unknown (Conway, 2020; Frost et al., 2015). There are mixed findings with a few patients with memory disorders that may reflect either variation in SL paradigms, patient abilities, or both (Schapiro et al., 2014; Covington et al., 2018; Cerrata et al, 2019; Dienes et al., 2021). Neuroimaging findings have revealed basal ganglia involvement across both SL and procedural learning tasks (McNealy et al., 2006; Karuza et al., 2013; Willingham et al., 2002; see Conway & Pisoni, 2008 for a review), while the hippocampus has also been shown sensitive to visual input structures (Schapiro et al., 2012; Tang et al., 2022; Wammes et al., 2022).

In the current study, we used a classic embedded-pattern learning paradigms (Saffran et al., 1996) to define and measure SL performance for the following reasons. First, we aimed to compare our findings with decades of empirical proof that typical adults are capable of robust SL across sensory modalities (e.g., visual shape and color sequences: Turk-Browne et al., 2008; auditory tones: Saffran et al., 1999; speech syllables: Saffran et al., 1996; see Frost et al., 2015 for a review on domain-generality vs. modality-specificity debate of SL). Second, successful embedded-pattern learning does not *require* explicit knowledge of the task goal or any motor engagement (Song et al., 2007; Batterink et al., 2015), which enabled us to test the domain-general procedural deficit hypothesis across SL and procedural learning tasks and assess the similarity and difference in performance.

Spoken and written language inputs are rich in regularities. Therefore, SL has been proposed as an important mechanism underlying typical language and reading development (Aslin & Newport, 2008; Erickson & Thiesoon, 2015; Romberg & Saffran, 2010; Arciuli, 2018; Sawi & Rueckl, 2019). Empirical evidence has tied SL with reading skills in both first and second language (Arciuli & Simpson, 2012; Spencer et al., 2015; Qi et al., 2019; Tong et al., 2019; Frost et al., 2013; Yu et al., 2019). When both auditory and visual SL were examined in typically reading adults and children as measured both by increasing speed and two-alternative forced choice accuracy (2AFC), reading skills were more strongly associated with auditory SL than visual SL (Qi et al., 2019). In children, the relationship between auditory SL and reading skills was further mediated by an emergent literacy skill: phonological awareness, suggesting implicit sequential learning, specifically in the auditory modality, might constitute the earliest steps towards phonological awareness development, a pivotal building block of literacy development.

There have been mixed findings of deficits in SL in individuals with dyslexia. In the visual modality, some studies reported similar learning patterns between dyslexic and typically reading individuals (van Witteloostuijn et al., 2021; Singh et al., 2018; Nigro et al., 2015; Howard et al., 2006), while others have reported impaired SL learning in dyslexia (Sigurdardottir et al., 2017; Tong et al., 2019). In the auditory modality, however, findings are more consistent, especially in adult participants. Across both linguistic and nonlinguistic stimuli, dyslexic adults appear to show less success in recognizing embedded auditory patterns (Gabay et al., 2015; Dobó et al., 2021; also see Singh & Conway, 2021 for a review). The lack of consensus in the literature regarding the status of SL in dyslexia is consistent with the pluralist view of SL (Frost et al., 2019) positing that SL across modalities and domains operates through partially overlapping, but distinct mechanisms. Therefore, a direct comparison between similarly designed auditory and visual SL tasks is necessary to reconcile whether certain types of SL are indeed more vulnerable than others in dyslexia.

The present study had two major aims. First, we asked whether adults with dyslexia would show intact or impaired procedural memory on two motor skill learning tasks that have been well established as reflecting purely procedural memory, mirror tracing and rotary pursuit. Intact learning in dyslexia would contradict the idea that there is a broad impairment of procedural memory in dyslexia. Second, given that reading development is built upon inputs from both the visual and auditory sensory modalities, we asked whether statistical learning, an implicit learning process, could be variably intact and impaired in dyslexia.

**Methods**

**Participants**

Twenty-six adults with dyslexia (16 female) and 27 typical readers (14 female) matched on age, sex ratio, and IQ (age 18-41 years, *M* = 26.6, *SD* = 6.3) participated in this study. All participants met eligibility criteria: being a native speaker of American English; born after at least 36 weeks’ gestation; no sensory or perceptual difficulties other than corrected vision; no history of head or brain injury or trauma; no neurological, neuropsychological, or developmental disorder diagnoses; no medications affecting the nervous system; nonverbal IQ standard score > 85 (Matrices subtest of the Kaufman Brief Intelligence Test/KBIT-2; Kaufman & Kaufman, 2004). Hearing tests were completed for all participants and participants with atypical hearing were excluded. The study was approved by the Committee on the Use of Humans as Experimental Subjects (COUHES) at MIT.

**Neuropsychological characterization**

All participants completed a comprehensive battery of standardized reading, language, and cognitive assessments, as well as a background questionnaire (**Table 1**). Measures included: Sight Word Efficiency (SWE) and Phonemic Decoding Efficiency (PDE) subtests of Test of Word Reading Efficiency (Torgesen, Wagner, & Rashotte, 2012); Word ID (WID) and Word Attack (WA) subtests of Woodcock Reading Mastery Test-Revised/Normative Update (Woodcock, 2011); Peabody Picture Vocabulary Test (Vocabulary; Dunn, Dunn, Lenhard, Lenhard, Suggate, 2015); Wechsler Adult Intelligence Scale (Adult-IQ; Wechsler, 2008-for adults); Elision and Blending subtests of Comprehensive Test of Phonological Processing (Wagner, Torgesen, & Rashotte, 1999); ​​and Digit Span subtest of the Wechsler Adult Intelligence Scale, 4th Edition was used to assess short term memory (WAIS-IV, Wechsler, 2008). Participants were included in the *developmental dyslexia* group (DD) based on performance below the 25th percentile on at least two out of four standardized subtests of timed or untimed word or nonword reading (SWE, PDE, WID, and WA). Participants were included in the *typical reader* group (TYP) based on performance at or above the 25th percentile on all four of the above subtests. The sample’s demographic information is available in Supplemental Table 2. A majority of participants in the *DD* group also reported an external diagnosis of dyslexia (*N* = 20) and a history of reading delay (*N* = 19). One participant in each group had a diagnosis of ADHD.

**Table 1**. Summary Behavioral Characterization of Participants

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Dyslexia (N=26) | Typical (N=27) | *p* value | *Cohen’s d* |
| Age | 26.35 (6.59) | 26.68 (6.58) | 0.848 | 0.18 |
| IQ | 107.30 (14.84) | 113.68 (13.97) | 0.106 | 0.28 |
| WID | 89.44 (9.54) | 110.75 (7.40) | < 0.001 | 2.54 |
| WA | 77.59 (9.71) | 102.46 (7.97) | < 0.001 | 2.81 |
| SWE | 89.19 (9.19) | 110.25 (12.50) | < 0.001 | 1.89 |
| PDE | 83.52 (8.19) | 106.54 (7.78) | < 0.001 | 2.87 |
| Vocabulary | 106.15 (10.02) | 113.93 (8.51) | 0.004 | 0.85 |
| Elision | 8.22 (2.31) | 9.89 (1.91) | 0.005 | 0.8 |
| Blending Words | 10.33 (3.16) | 12.50 (2.69) | 0.008 | 0.75 |
| Digits Span | 8.63 (2.37) | 11.04 (2.38) | < 0.001 | 1.01 |

**Tasks**

***Mirror Tracing***

Participants *(N:* DD = 26, TYP = 27) watched their hands in the mirror while tracing the outline of a six-sided star (Gabrieli et al., 1993; Milner, 1962). A Lafayette Instruments Auto-Scoring Mirror Tracer, a device that includes a metal stylus and a metal test plate (except for the star pattern), was used for this task (Model 58024A⁄C). When the stylus goes off the star and touches the metal plate, it completes an electrical circuit and an error is recorded. Participants were instructed to trace as quickly and accurately as possible, while staying inside the outline of the star. Participants first completed a practice trial and then traced four times. After 30 min of performing other tasks, participants traced five more times. Completion time and number of errors per trial were the dependent variables.

***Rotary Pursuit***

Participants *(N:* DD = 26, TYP = 27) used a Lafayette photoelectric pursuit rotor (Lafayette Instruments, Model 30014) to maintain contact with a photoelectric target that rotated in the shape of a rectangle with truncated corners using a stylus. Participants first completed a 20-sec practice trial to establish baseline speed (15, 30, 45, or 60 rotations per minute). The speed at which a participant’s time-on-target was closest to 5 sec was selected as the baseline and used for all subsequent trials. Participants then completed four 20-sec trials, took a break for 1 min, and then completed four more 20-sec trials. After 30 min of performing other tasks, participants completed eight more 20-sec trials, taking a 1-min break after the first four trials as before. The dependent measure was the proportion of time-on-target divided by time-off-target per trial.

***Statistical Learning (SL)***

Participants (*N*: DD = 17, TYP = 24) completed one visual SL (VSL) task and one auditory SL (ASL) task, hosted on a secured website (<https://www.cogscigame.co>). The detailed design and the procedure of the SL tasks have been previously described in Qi et al. (2019). For each SL task, a familiarization phase, in which participants performed a target detection task for about 5 min, was immediately followed by a test phase, in which a two-alternative forced-choice (2-AFC) test was given.

In the familiarization phase, stimuli were presented in a continuous stream according to an embedded pattern of four unique triplets. In VSL, 12 unique alien images formed four target triplets. Each of the target triplets was repeated 24 times for a total of 96 triplets. Each image was presented one at a time at the center of the screen for 800 ms with 200 ms of inter-stimulus interval (stimulus onset asynchrony (SOA) = 1000 ms), lasting 4 min 48 sec. Participants were instructed to press the spacebar as quickly as possible whenever the target alien appeared on the screen. The target alien image was always the third alien of one of the four base triplets so that online learning could be measured via response time acceleration over 24 target trials during exposure. In ASL, 12 unique monotones of the same duration (328 ms) (Tone Task) formed four target triplets. Each triplet was repeated 48 times for a total of 192 triplets. The SOA was 480 ms, with the familiarization phase lasting 4 min and 36 sec. Presentation speed was faster in the auditory than visual tasks due to differences in perceptual preference (Conway & Christiansen, 2009; Emberson, Conway, & Christiansen, 2011). The procedure was identical to that of VSL except that the target tones used in the target-detection task during familiarization were constrained to only the lowest and highest notes of the final tones of the four triplets to facilitate identification. Two practice trials before the continuous stream of tones ensured that participants could distinguish the target tone. Response time was measured over 48 target trials. This approach of measuring online learning has been validated in our previous work in adult learners who accelerated more quickly in their responses to target stimuli in structured sequences, similar to the ones used here, than the target stimuli in random sequences where no triplets were formed and the same stimuli were displayed in a random order (Schneider et al., 2020; Tang et al., 2022). Significant RT acceleration was also observed in children using a tablet to respond to target stimuli at the final position of a triplet during the exposure phase of SL, but not for the target stimuli at the start position of a triplet (Zinszer et al., 2020).

The 2AFC test phase taps into post-learning reflection of the learned patterns. The task was introduced after the familiarization phase and was composed of 32 questions. For each question, two options were included: a target triplet from the familiarization phase and a foil triplet that was novel to the participant. Foil triplets were constructed so that the relative position of each image in the foil triplet was the same as the target triplet. The test phase consisted of 32 (4 target triplets x 4 foil triplets x 2 repetitions) randomly ordered trials. The images and sounds within each triplet were presented one at a time at the same presentation rate as the familiarization phase with a 1000 ms pause between the target and the foil triplets. Participants were asked to identify which of the two triplets (embedded versus foil) seemed more like what they saw during the familiarization phase. There were no time constraints for responses and no feedback on the accuracy of answers. The order of test trials was randomized across participants.

One participant in the DDgroup was removed from the ASL analyses and one participant in the TYP group was removed from the VSL analyses because their data were not recorded due to technical issues. Task order (VSL or ASL first) was counterbalanced across participants.

**Statistical Analysis**

For all four experiments, outcome variables were analyzed in R v3.5.0 (R Core Team, 2013), using identical statistical thresholds (*p* < 0.05*)*, and maximal random effect structures (Barr et al., 2013), using the package *lme4* (Bates et al., 2007). The significance of fixed effects in the linear mixed-level models was tested in an ANOVA (using Satterthwaite approximations) and fit with restricted maximum likelihood (REML) using the package *lmerTest* (Kuznetsova et al., 2016). The significance of fixed effects in the generalized linear mixed-level models was estimated using personalized quasi-likelihood using the *glmmPQL* function of the package *MASS* to improve model convergence. The effect sizes for the fixed effects in these linear mixed models were computed via the coefficient of determination (R squared) using *r2beta* function of the package *r2glmm* (Edwards et al., 2008; Nakagawa & Schiezeth, 2013; Jaeger et al., 2016). In all models, Sex, Age, and nonverbal IQ were included as nuisance covariates.

***​​Pairwise Correlation Analysis***

To test whether individuals’ performance across these different tasks is partially constrained by an underlying unified capacity, Pearson pairwise correlations were computed to test for cross-task associations using the Hmisc package in R (Harrell Jr & Harrell Jr, 2019). Total completion time across trials was extracted for the MT and RP tasks, and mean number of errors were extracted for the MT task. Each participant’s SL performance was measured by 1) the linear slope of RT acceleration over normalized response time, so that we are able to compare SL performance across individuals with different baseline speed, and 2) the proportion of correct response during the 2AFC task. Bayesian correlations were computed using the BayesFactor package (Morey et al., 2015) with default priors comparing a null model of no correlation with the alternative model of correlation. Bayesian models provide good precision even in smaller data sets (Lee & Song, 2004). Importantly, Bayes factors provide a measure of how likely the data is under the null versus alternative hypothesis, allowing us to quantify and compare relative support for the existence of a relationship between each pair of variables. Based on the previous work, Bayes factors larger than 1 were considered to provide positive evidence (albeit weak if under 3) in favor of the alternative hypothesis that two variables are correlated (Jeffreys, 1998; Wetzels et al., 2011).

**Results**

There were no significant group differences in age or IQ, but participants with dyslexia performed significantly worse than typical readers on phonological awareness, working memory, and vocabulary measures.

**Rotary Pursuit**

The group performance by trial is shown in **Figure 1A**. There were no significant group differences in the baseline speed (*t*(45.39) = 0.06, *p* = 0.95, d = 0.02). Both groups showed substantial improvement across trials. To test for group differences in time on target, a linear mixed-effects model was conducted with *proportion on* (computed as time on target/ (time on target + time off target)) as the dependent variable. Fixed factors in the model included *trial number* and *group* (DD vs. TYP); the model’s random effects structure included random intercepts by participants and by-participant random slopes on trial (Jaeger, 2008). The main effect of *trial* was significant (*b* = 0.01, SE = 0.002, *t* = 7.13, *p* < 0.001, 𝑅2𝑚 = 0.072), with an increase in the proportion of time on target across trials for both groups. The main effect of *group* was not significant (*b* = 0.01, SE = 0.04, *t* = 0.30, *p* = 0.77, 𝑅2𝑚 = 0). The interaction between *trial number* and *group* was marginal (*b* = -0.005, SE = 0.002, *t* = -1.86, *p* = 0.069, 𝑅2𝑚 = 0.005), suggesting a marginally steeper slope in the DD group (*b* = 0.01, SE = 0.001, *t* = 8.37, *p* < 0.001) than the TYP group (*b* = 0.008, SE = 0.002, *t* = 4.15, *p* < 0.001).

**Mirror Tracing**

The completion time by trial and the number of errors by trial are shown in **Figure 1B** and **Figure 1C** respectively. There were no significant group differences in the baseline time (*t*(24.60) = 0.46, *p* = 0.65, d = 0.14) or error (*t*(38.23) = 0.05, *p* = 0.96, d = 0.01) during the first practice trial. Both groups showed substantial reduction in completion time and number of errors across trials. To test for group differences, two linear mixed-effects models were conducted with completion *time* and total number of *errors* as the dependent variables. Fixed effects in the models included *trial number* (Trials 2–10 because the first trial was practice) and *group* (DD vs. TYP); the model’s random effects structure included random intercepts by participants and by-participant random slopes on trial. Age, IQ, and Sex were included as nuisance covariates.

For the *time* model, the main effect of *trial* was significant (*b* = -2.43, SE = 0.68, *t* = 3.59, *p* = 0.001, 𝑅2𝑚 = 0.043), with a reduced time on task across trials. Neither the main effect of *group* (*b* = 3.21, SE = 10.20, *t* = 0.32, *p* = 0.75, 𝑅2𝑚 = 0.001) nor the interaction between *trial* and *group* (*b* = 0.14, SE = 0.94, *t* = 0.15, *p* = 0.88, 𝑅2𝑚 = 0) weresignificant. Similar results were revealed for the *error* model. The main effect of *trial* was significant (*b* = 1.88, SE = 0.43, *t* = 4.32, *p* < 0.001, 𝑅2𝑚 = 0.058), with a reduced number of errors across trials. Neither the main effect of *group* (*b* = 6.98, SE = 5.84, *t* = 1.20, *p* = 0.24, 𝑅2𝑚 = 0.01) nor the interaction between *trial* and *group* (*b* = 0.71, SE = 0.61, *t* = 1.17, *p* = 0.25, 𝑅2𝑚 = 0.004) weresignificant.

**Auditory Statistical Learning**

Familiarization Phase: Valid button presses for the target tones were defined from the onset of the previous stimulus to the onset of the next two stimuli (-480 ms - 960 ms) to allow for anticipatory button presses as well as delayed button presses for the target tones. Eight participants (2 in the DD group and 6 in the TYP group) were removed from the analyses due to fewer than 25% of the valid responses to a total of 48 target stimuli presentations. As a result, 14 participants in the DD group and 18 participants in the TYP group are included in the reaction time analyses. The two groups were not significantly different in their baseline response time (*t*(25.9) = 1.52, *p* = 0.14, *d* = 0.57) or hit rate (*t*(30.0) = -1.38, *p* = 0.18, *d* = 0.49; Supplementary Table 1).

Graphical user interface, histogram

Description automatically generated

**Figure 1**. The time courses of procedural learning and implicit SL during the familiarization phase. Shown is performance for typical readers (TYP, dash-dotted lines) and adults with dyslexia (DD, solid lines) plotted across trials for rotary pursuit (A), mirror tracing (B, C), the familiarization phase of auditory statistical learning (D), and the familiarization phase of visual statistical learning (E). The vertical dash-dotted lines in 1A-1C indicate task breaks.

The group performance by trial is shown in **Figure 1D**. Neither group showed a trend of acceleration over the course of familiarization. To test for group differences, a linear mixed-effect model was conducted with reaction *time* as the dependent variables. Fixed effects in the models included *trial number* (Trials 1–48), *group* (DD vs. TYP), and their interaction. Age, gender, and non-verbal IQ were included as covariates. The model’s random effects structure included random intercepts by participants and by-participant random slopes on *trial number*. There was no significant effect of *trial number* (*b* = 0.20, SE = 1.48, *t* = 0.14, *p* = 0.89, 𝑅2𝑚 = 0), *group* (*b* = 67.22, SE = 60.88, *t* = 1.10, *p* = 0.28, 𝑅2𝑚 = 0.004), or their interaction (*b* = 0.24, SE = 1.92, *t* = 0.13, *p* = 0.90, 𝑅2𝑚 = 0) on reaction time.

Test Phase: Participants’ overall accuracy in the 32-trial 2AFC test was depicted in **Figure 2**. Both groups performed significantly above 50% chance level (DD: Mean = 0.55, SD= 0.09, *t*(15) = 2.19, *p* = 0.02; TYP: Mean = 0.66, SD= 0.13, *t*(23) = 6.22, *p* < 0.001). We compared the two groups using a generalized linear mixed model fit by maximum likelihood (Laplace Approximation). The dependent variables are each participants’ trial-by-trial binomial accuracy data. The fixed effect included *group* (DD vs. TYP). Age, gender, and non-verbal IQ were included as covariates. The model’s random effects structure included random intercepts by participants and by trial. There was a significant main effect of the group. The DD group had significantly lower accuracy in identifying the target tone triplets from the foil triplets compared to the TYP group (*b* = 2.43, SE = 0.65, *z* = 3.71, *p* < 0.001, 𝑅2𝑚 = 0.02).

**Visual Statistical Learning**

Familiarization Phase: Valid button presses for the target tones were defined from the onset of the previous stimulus to the onset of the next two stimuli (-1000 ms - 2000 ms) in order to allow for anticipatory button presses as well as delayed button presses for the target tones. Participants performed the target detection task with high hit rates. As a result, all participants were kept in this analysis. The two groups are not significantly different in their baseline response time (*t*(34.5) = 0.70, *p* = 0.49, *d* = 0.23) or hit rate (*t*(21.7) = -0.70, *p* = 0.13, *d* = 0.57; Supplementary Table 1).

The group performance by trial is shown in **Figure 1E**. To test for group differences in response time changes over the course of familiarization, a linear mixed-effect model was conducted with reaction *time* as the dependent variables. Fixed effects in the models included *trial number* (Trials 1–24), *group* (DD vs. TYP), and their interaction. The model’s random effects structure included random intercepts by participants and by-participant random slopes on *trial number*. There was a significant effect of *trial number* (*b* = 2.56, SE = 0.91, *t* = 2.81, *p* = 0.008, 𝑅2𝑚 = 0.01) and a marginal interaction between *group* and *trial number* (*b* = 2.10, SE = 1.19, *t* = 1.76, *p* = 0.087, 𝑅2𝑚 = 0.004) on reaction time. The *group* difference on reaction time was not significant(*b* = -8.14, SE = 24.39, *t* = -0.33, *p* = 0.74, 𝑅2𝑚 = 0). Post-hoc within-group analyses suggest only the DD group showed a significant acceleration over the course of learning (*b* = -2.50, SE = 1.08, *t* = 2.31, *p* = 0.035, 𝑅2𝑚 = 0.02), but the TYP group (*b* = -0.46, SE = 0.65, *t* = 0.71, *p* = 0.49, 𝑅2𝑚 = 0.005) did not show significant acceleration during the familiarization phase.

Test Phase: Participants’ overall accuracy in the 32-trial 2AFC test was depicted in **Figure 2**. Both groups performed significantly above 50% chance level (DD: Mean = 0.72, SD= 0.21, *t*(16) = 4.39, *p* < 0.001; TYP: Mean = 0.67, SD= 0.23, *t*(22) = 3.55, *p* < 0.001). We compared the two groups using a generalized linear mixed model. The dependent variables are each participants’ trial-by-trial binomial accuracy data. The fixed effect included *group* (DD vs. TYP). *age, gender,* and *non-verbal IQ* were included as covariates. The model’s random effects structure included random intercepts by participants and by trial. The two groups were not significantly different on the 2AFC performance (*b* = 0.75, SE = 1.67, *t* = 0.45, *p* = 0.65, 𝑅2𝑚 = 0.001).

**Comparison of Auditory Statistical Learning and Visual Statistical Learning**

Familiarization Phase: We compared the group differences in RT changes across the two SL tasks. The online learning improvements in RT over the course of VSL were marginally larger in VSL than in ASL (three-way interaction between trials, tasks, and groups: *b* = 0.02, SE = 0.01, *t* = 1.87, *p* = 0.062, 𝑅2𝑚 = 0.002), suggesting a specific advantage in online VSL in DD.

Test Phase: We found a significant difference in learning between the groups on the two SL tasks: the group difference (TYP > DD) was significantly greater in the ASL task than the VSL task (*b* = 3.08, SE = 0.86, *t* = 3.59, *p* < 0.001, 𝑅2𝑚 = 0.005). This directly supports a dissociation in DD between reduced ASL learning and preserved VSL learning.

Chart, box and whisker chart

Description automatically generated

**\*\*\***

**\*\*\***

***ns***

**Figure 2**. Implicit SL performance during the test phase. Mean proportion of correct trials in adults with dyslexia (DD, darker grey) and typical readers (TYP, lighter grey) for auditory statistical learning and visual statistical learning. \*\*\*, *p* < 0.001.

**Pairwise Task Correlation Analysis**

All task performance measures presented moderate-to-good internal consistency as measured by Cronbach’s Alpha (Supplementary Table 3). To test whether our learning measures represent separate or overlapping skills, we examined the Pearson pairwise correlations across all seven measures. Our results revealed no significant associations among the different learning tasks. This was confirmed using Bayes factors with no evidence against the hypothesis for cross-task associations. The only significant associations were between MT accuracy and response time and between VSL accuracy and response time. The former correlation represents a tradeoff between accuracy and response time. The latter relationship is consistent with previous research (Qi et al., 2019), suggesting quicker RT acceleration during exposure was associated with greater success in recognizing the learned triplets.

**Table 2: Pairwise learning task correlations**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **1** | **2** | **3** | **4** | **5** | **6** |
| **1. RP Mean Prop On** | 1 |  |  |  |  |  |
| **2. MT Time** | 0.186  BF = 0.399 | 1 |  |  |  |  |
| **3. MT Error** | 0.104  BF = 0.224 | 0.368\*  BF= 5.407 | 1 |  |  |  |
| **4. ASL Accuracy** | -0.014  BF = 0.205 | -0.052  BF = 0.399 | 0.132  BF = 0.274 | 1 |  |  |
| **5. VSL Accuracy** | -0.01  BF = 0.205 | -0.028  BF = 0.207 | 0.0431  BF= 0.211 | -0.116  BF= 0.253 | 1 |  |
| **6. ASL RT Slope** | 0.221  BF = 0.439 | -0.028  BF = 0.229 | -0.223  BF= 0.443 | -0.034  BF= 0.223 | -0.230  BF= 0.469 | 1 |
| **7. VSL RT Slope** | -0.265  BF = 0.684 | -0.104  BF = 0.245 | -0.165  BF = 0.253 | 0.081  BF = 0.224 | -0.586\*\*  BF = 400.962 | -0.100  BF = 0.256 |

\*p < 0.05, \*\* p < 0.001

**Relationship between learning and standardized testing of reading and phonological skills.**

To examine the relationships between learning performance and individuals’ reading abilities, we assessed the Pearson correlations between all seven learning measures and the average of Word ID and Word Attack from WRMT-R (Table 3). We chose the untimed decoding skills because three out of seven learning measures involve response time and may inflate the correlations. Greater decoding skills were significantly associated with greater ASL accuracy (Figure 3A; *R* = 0.49, one-tailed *p* = 0.001, Bonferroni-corrected *p* < 0.05, BF = 24.706), but not with VSL accuracy (R = -0.12, one-tailed *p* = 0.27, BF = 0.452). The ASL-decoding relationship was significant within the DD group alone (*R* = 0.50, one-tailed *p* = 0.03, BF = 2.04), suggesting the significant association in the whole sample was not driven by the group difference.

We also examined the correlations between learning performance and individuals’ phonological awareness skills, measured by the average of Elision and Blending Words from CTOPP-2 (Table 3). A greater VSL RT slope in the DD group correlated significantly with higher phonological awareness scores (Figure 3B; *R* = -0.85, one-tailed *p* < 0.001, Bonferroni-corrected *p* < 0.05, BF = 244.077). There was also a significant correlation between greater VSL accuracy and higher phonological awareness scores in the DD group (*R* = 0.54, one-tailed *p* = 0.02, BF = 2.853). No correlations survived corrections for multiple comparisons within the typical group, perhaps due to a smaller variance in the reading and phonological awareness measures.

**Table 3: Correlation coefficients between procedural/statistical learning and standardized reading and phonological test scores.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Learning | Decoding | | Phonological Awareness | | | |
| All | DD | TYP | All | DD | TYP |
| RP Mean Prop On | -0.11  BF = 0.23 | 0  BF = 0.24 | -0.07  BF = 0.25 | 0.15  BF = 0.3 | 0.10  BF = 0.28 | 0.33  BF = 0.94 |
| MT Time | 0  BF = 0.18 | 0.16  BF = 0.33 | 0.28  BF = 0.56 | 0.07  BF = 0.18 | 0.1  BF = 0.28 | 0.16  BF = 0.32 |
| MT Error | 0.14  BF = 0.17 | 0.14  BF = 0.3 | 0.31  BF =0.79 | 0.05  BF = 0.17 | 0.01  BF = 0.24 | 0.22  BF = 0.46 |
| ASL Accuracy | 0.488\*\*  BF = 21.37 | 0.50\*  BF = 1.7 | 0.20  BF = 0.38 | 0.26  BF = 0.21 | 0.21  BF = 0.41 | 0.01  BF = 0.26 |
| VSL Accuracy | -0.12  BF = 0.26 | -0.31  BF = 0.58 | 0.08  BF = 0.28 | 0.23  BF = 0.21 | 0.54\*  BF = 2.54 | 0.19  BF = 0.38 |
| ASL RT Slope | 0.08  BF = 0.34 | 0.02  BF = 0.34 | 0.12  BF = 0.33 | 0.02  BF = 0.23 | 0.35  BF = 0.64 | 0.04  BF = 0.38 |
| VSL RT Slope | 0.17  BF = 0.25 | 0.04  BF = 0.31 | -0.09  BF = 0.29 | -0.01  BF = 0.69 | -0.71\*\*\*  BF = 24.46 | 0.14  BF = 0.9 |

Note: \* uncorrected p < 0.05; \*\* uncorrected p < 0.01; \*\*\* uncorrected p < 0.001. Underlined values represent significant correlations after Bonferroni corrections for fourteen correlations within each group.



**Figure 3. Relationships between statistical learning and reading-related skills: A. untimed word and nonword decoding skills. B. Phonological Awareness.**

**Discussion**

According to an influential theoretical framework, reading deficits in developmental dyslexia result from atypical domain-general procedural learning (Nicolson & Fawcett, 2011; Ullman et al., 2020; Ullman & Pullman, 2015). Here we tested this theory using four non-linguistic tasks that measure implicit learning in adults with and without dyslexia. We found no evidence for impaired learning in adults with dyslexia on rotary pursuit and mirror tracing, two classical and well-validated paradigms designed to measure procedural learning independent of declarative memory. In the auditory statistical learning task, individuals with dyslexia showed reduced auditory statistical learning as measured by test accuracy in post-learning triplet recognition. In contrast, individuals with dyslexia showed typical learning on the visual statistical learning task. Thus, these findings suggest that instead of a pervasive deficit in procedural learning, learning deficits in dyslexic adults are specific to the domain that shows the most consistent impairment in dyslexia: auditory processing.

The findings that adults with dyslexia show equivalent learning to that of typical readers on the two motor learning tasks challenge previous reports of broad procedural deficits in dyslexia (e.g., Lum, Ulman, and Conti-Ramsden, 2013; Nicolson & Fawcett, 2007; 2011; Ullman, 2004; Ullman & Pierpont, 2005). There is ample evidence that skill learning on both mirror tracing and rotary pursuit tasks is strongly dissociable from declarative or explicit memory (e.g., Milner, 1962; Corkin, 1968; Gabrieli, Corkin, Mickel, & Growdon, 1993; Heindel, Salmon, Shults, Walicke, & Butters, 1989). The absence of differences in learning between the typical and dyslexia group in the current study combined with lack of convincing evidence for procedural deficits in dyslexia reported in two recent meta-analyses (West et al., 2021; Oliveira et al., 2022), point convincingly against broad deficits in procedural learning as the core deficit in dyslexia.

The statistical learning results further rule out a domain-general account of procedural learning deficit. We found that dyslexic adults showed specific impairment in ASL, but intact ability in VSL. Moreover, reading skills, measured by decoding, were strongly associated with ASL, but not with VSL, both across the entire sample and within the dyslexic group. The specificity of the association between reading and ASL in the dyslexia group, but not in typical readers may suggest underlying low-level perceptual deficits that impair performance on both tasks. These findings are consistent with the specific link between ASL and reading skills reported before in neurotypical adults and children (Qi et al., 2019). Atypical auditory learning might be one of the underlying causes for challenges faced by dyslexic individuals in establishing grapheme-to-phoneme mapping. Decades of dyslexia research have documented the widespread and persistent deficits in the speech sound domain not only in dyslexic adults, but also in at-risk pre-readers who later developed dyslexia (Wagner & Torgesen, 1987; Manis et al., 1996; Carroll & Snowling, 2004; Ramus & Szenkovits, 2008; Melby-Lervåg et al., 2012; Ramus et al., 2013). Longitudinal studies in typical children have provided some of the strongest evidence for a causal role of phonological skills in reading development, that is phonological skills and their neural correlates in pre-readers or beginning readers (e.g., Perfetti et al., 1987; Lervåg et al., 2009; Saygin et al., 2013; Wang et al, 2020), predict children’s later reading skills. It is possible that the delay in early phonological development snowballed into later reading difficulties in dyslexia.

Reduced ASL might reflect differences in rapid auditory processing in dyslexia (Ahissar et al., 2000; Heath & Hogben, 2004; Tallal, 1980). Individuals with dyslexia demonstrated difficulties in tracking, or anchoring to, repeated auditory stimuli such as tones (Ben-Yehudah, Banai, & Ahissar, 2004; Amitay et al., 2002; Ahissar et al., 2006), and voice recognition (Perrachione et al., 2011). These deficits may be the result of less reliable auditory perception rather than learning/adaptation deficits (Ozernov-Palchik et al., 2021). For example, adults with dyslexia demonstrated reduced capacity for tone-frequency discrimination but benefitted to the same extent in their performance as typical readers when one comparison tone was held constant across trials (Ozernov-Palchik et al., 2021). Other studies, however, have demonstrated deficits in adaptation even when baseline auditory discrimination skills were matched across the groups (Gabay & Holt, 2021; Gabay et al., 2022).

Our study is in part consistent with prior SL findings in the dyslexic literature. To date, the majority of SL studies in dyslexic adults and adolescents have only investigated a single sensory modality. Yet less efficient learning in ASL (Gabay et al., 2015; Kahta & Schiff, 2019; Dobó et al., 2021), as well as in VSL (Sigurdardottir et al., 2017; Kahta & Schiff, 2016), have both been documented. Notably, the impairment in SL seems to hinge on the implicit nature of the task, because no behavioral difference was found when dyslexic adults were either informed of the embedded statistical patterns prior to learning or became vaguely aware of the embedded patterns after learning (Kahta & Schiff, 2016; Sigurdardottir et al., 2017). The visual saliency of our alien cartoon stimuli, together with our target-detection cover task, may have boosted attention to the stimuli and therefore facilitated learning (Toro et al., 2005; Turk-Browne et al., 2005; Schneider et al., 2022).

The positive relationship between visual statistical learning and phonological awareness in dyslexic adults was interesting, and somehow unexpected. This relationship was not found in a typically reading population (Qi et al., 2019). However, a reversed causal relationship between phonological skills and reading has been proposed in typical reading development as well, that is, older school-aged children improve phonological abilities through their reading experiences (e.g., Castles & Coltheart, 2004). Similarly, dyslexic adults might also hone their phonological skills through reading and decoding practices and superior visual statistical learning might boost the benefits. This compensatory route might be especially valuable for learners whose spoken language and written language skills develop in tandem. This possibility has been recently supported by a study in a group of beginning readers of a second language. Children’s visual statistical learning was found to predict their phonological awareness in the second language (Zinszer et al., 2022).

Our study has a few methodological limitations. First, the mean effect size reported for group differences on procedural learning tasks in adult studies is small (g = 0.23; West et al., 2022), and a much larger sample size per group than the one employed in the current study would be needed to detect such small differences (~N = 297 participants per group). In contrast to the small effect sizes for procedural learning deficits, however, group differences in phonemic awareness yielded much large effect sizes across studies (d = 1.37; Melby-Lervåg, Lyster & Hulme, 2012) and in the current study (d = 0.88). Therefore, although our study was potentially underpowered to detect procedural learning differences between groups, if deficits in procedural learning were important causal factors in dyslexia, larger effect sizes for group differences would be expected (such as the ones for phonological awareness). Second, as indicated by a much lower hit rate, the target detection cover task in ASL was a much more difficult task compared to VSL. As a result, this measure can be heavily influenced by individual differences in perceptual acuity for tones or attention during learning. During ASL, neither group showed any evidence of RT acceleration. Previous studies observed similar null results at the group level in ASL (e.g., Qi et al., 2019; Schneider et al., 2020), but individual differences in ASL RT acceleration can still serve as a valuable predictor for reading-related skills. For example, neurotypical children’s ASL RT slope was significantly related to decoding skills, which was mediated by phonological awareness (Qi et al., 2019). Third, our SL measures, despite capturing both online and offline learning, are not sufficient to tease apart learning and retrieval mechanisms. The above-chance 2AFC accuracy does not depend on RT acceleration, nor is RT acceleration solely driven by pattern learning. Future research with neuroimaging approaches is necessary to pinpoint which subprocess of learning is more vulnerable in dyslexia.

In conclusion, our study combining four classic procedural learning and statistical learning tasks provides converging evidence against the domain-general procedural learning deficit in dyslexia adults. Even though a shared subcortical contribution to procedural learning across all four tasks is well-documented (Janacsek et al., 2022), dyslexic adults show reduced performance only in auditory statistical learning, but typical, and even slightly better, performance in motor skill learning and visual statistical learning. Difficulties in learning phoneme-to-grapheme mapping in dyslexia, therefore, cannot be directly attributed to the procedural dysfunctions governed by the core subcortical circuitry involving basal ganglia (Krishnan et al., 2016; *c.f.* Ullman et al., 2020). Instead, our findings suggest reading acquisition in dyslexic individuals might be constrained specifically by neural substrates of auditory processing/learning, providing support for a multi-component and pluralist view of learning (e.g., Frost et al., 2019; Bogaerts et al., 2022).

**Supplementary Table 1**: Performance during the SL familiarization phase.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | ASL | | VSL | |
| Mean (SD) | Mean RT (ms) | Hit rate | Mean RT (ms) | Hit rate |
| DD | 408.6 (102.6) | 0.55 (0.20) | 475.4 (70.8) | 0.96 (0.06) |
| TYP | 356.0 (88.76) | 0.66 (0.25) | 491.2 (70.1) | 0.99 (0.03) |

**Supplementary Table 2**= Demographic information

|  |  |  |
| --- | --- | --- |
|  | DD (N=26) | TYP (N=27) |
| History of Reading Delay | 19 (73.1%) | 1 (3.7%) |
| History of Language Delay | 2 (7.7%) | 0 (0.0%) |
| Diagnosis Dyslexia | 20 (76.9%) | 0 (0.0%) |
| Diagnosis ADHD | 1 (3.8%) | 1 (3.7%) |
| Race |  |  |
| White | 22 (84.6%) | 24 (88.9%) |
| Black/African American | 4 (15.4%) | 3 (11.1%) |
| Hispanic | 0 (0.0%) | 1 (3.7%) |
| Income |  |  |
| <30k | 6 (24.0%) | 3 (11.5%) |
| 30-60k | 11 (44.0%) | 9 (34.6%) |
| 60-100k | 4 (16.0%) | 3 (11.5%) |
| >100k | 4 (16.0%) | 11 (42.3%) |
| Education |  |  |
| Less7Grd | 0 (0.0%) | 0 (0.0%) |
| Junior High | 0 (0.0%) | 0 (0.0%) |
| High School | 1 (3.8%) | 0 (0.0%) |
| Partial College | 2 (7.7%) | 0 (0.0%) |
| College | 7 (26.9%) | 12 (44.4%) |
| Masters | 14 (53.8%) | 10 (37.0%) |
| Doctorate | 2 (7.7%) | 5 (18.5%) |

**Supplementary Table 3**: Cronbach’s alpha for all tasks.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Procedural Learning** | | | **Statistical Learning** | | | |
|  | Rotary Pursuit | Mirror Tracing - Error | Mirror Tracing - Time | VSL - RT | VSL - Accuracy | ASL - RT | ASL - Accuracy |
| Cronbach’s Alpha | 0.98 | 0.68 | 0.56 | 0.94 | 0.89 | 0.89 | 0.65 |

**References**

Arciuli, 2018, reading as statistical learning. Lang.Speech Hear. Service. Sch. 49, 634-643

Sawi & Rueckl (2019). Reading and the neurocognitive bases of statistical learning. Sci. Stud. Read. 23, 8-23.