Dear Dr. Kiran,

Thank you for giving us the opportunity to submit a revised draft of our manuscript, “Intact Procedural Memory and Impaired Auditory Statistical Learning in Adults with Dyslexia.”

We are very grateful to the reviewers for their timely review of the manuscript and their insightful comments. We have carefully addressed their comments and integrated their recommendations into the revised manuscript. These revisions are highlighted using tracked changes within the document, and our detailed responses to each point are provided below in blue. We believe that this review process has considerably enhanced the quality of our manuscript.

Best regards,

Ola Ozernov-Palchik and Zhenghan Qi

**Reviewer #1**

1. I know that it is not easy to operate with several different terms and that the demarcation lines are not clear between procedural, implicit and statistical learning. I see the attempts to introduce these terms in a logical way, but I think this still needs some work. In particular, I would remove the section of the historical description of declarative/procedural memory systems, I think that’s a bit unnecessary. Instead, I think it would be beneficial to explain and introduce the different terms at the beginning of the introduction, such as in the second paragraph, and describe how they are related to each other, then summarize the findings in dyslexia in later paragraphs.

We thank the reviewer for this important point. We have revised the Introduction to better clarify the relationships between the concepts of implicit learning, procedural memory, and statistical learning. To achieve this, we have defined ‘implicit learning’ as a comprehensive term for learning that occurs unconsciously and depends (at least in part) on the procedural memory system. Furthermore, we have distinguished two types of implicit learning: skill learning and statistical learning, and provided an overview of the paradigms used to investigate each type. We hope that these revisions will help readers better understand the concepts discussed in the paper.

1. Also, I found the title a bit misleading – I understand that most forms of procedural learning were intact in dyslexia, but right now, the impaired auditory statistical learning is not appearing in the title.

In accordance with Reviewer’s concern, the revised title is **Intact Procedural Memory and Impaired Auditory Statistical Learning in Adults with Dyslexia**

1. The statistical learning tasks were conducted on considerably less participants than the other two tasks. Why’s that? Relatedly, a high number of participants were excluded from the analyses of the auditory statistical learning paradigm. Why did you decide on the 25% threshold? Also, is the matching between the clinical and control groups still stands after excluding that many participants? Moreover, could you please report the results on the whole sample on the auditory statistical learning task?

Thank you for raising this important question about the sample size. We have now provided more information on the participants and SL tasks in our methods section. Due to the online nature of the SL tasks and participants completing them independently after their lab visit for the main study, not all participants chose to complete these tasks, resulting in a smaller sample size for these two tasks.

In this smaller subset of participants, age and gender ratio reflected the proportions of the larger group. Nonverbal IQ, however, was lower in the DD group than in the TYP group (*p* = 0.02). We have now described the nonverbal IQ differences in the methods on page 8. This group difference in nonverbal IQ, however, does not affect our conclusion because age, gender, and nonverbal IQ have been included as covariates for all of our group comparison models consistently throughout our manuscript.

Furthermore, for the ASL task, we examined the distribution of the hit rate across our sample. 0.25 threshold falls within 1.5 SD below the Mean. As both Reviewer 1 and Reviewer 2 pointed out, this threshold might be too stringent and has caused us to remove too many participants from the slope analyses. Therefore, for group comparison, we followed the Reviewer 1’s suggestion by including all participants into the RT analyses. The results stayed the same. But to avoid skewing our correlational analyses, we opted to remove one TYP participant who showed an unusual deceleration response over the course of exposure (RT slope was more than 3 SD above the sample mean) from the correlations that involve ASL RT slopes. These changes do not affect our main results, because all the ASL 2AFC analyses had already included every participant, regardless of their behavior during the exposure phase.

1. In the Comparison of Auditory Statistical Learning and Visual Statistical Learning section you state that “The online learning improvements in response time were marginally larger in VSL than in ASL (three-way interaction between trial number, task, and group: b = 0.02, SE = 0.01, t = 1.87, p = 0.062, 2 = 0.002), suggesting a specific advantage in VSL in DD.” This means that the difference in response time was only present in the DD group but not in the TYP group? Could you please report the exact statistics on how the tasks/groups differ?

We are sorry for any confusion. The statistics detailing the differences between the groups were provided in each SL task. The purpose of this section was to compare the degree of group differences between the two tasks. The marginal three-way interaction indicated that the group difference in VSL RT was greater than in ASL RT. However, Reviewer 2 recommended removing this sentence due to the marginal effect. We have updated this section to improve clarity.

Familiarization Phase: To compare the degree of group differences in response-time changes across the two SL tasks, we tested the interaction between trial number, task, and group in a linear mixed model. Our analysis revealed a marginal three-way interaction (b = 0.01, SE = 0.005, t = 1.84, p = 0.066, 𝑅2𝑚 = 0.002), that is, the group difference (DD quicker than TYP) in real-time VSL is marginally larger than the group difference in real-time ASL.

1. In Table 2 and 3, could you please report the exact p values? Also, when reporting Bayes Factors, I would write BF10, so it is clear that the evidence is for the alternative hypothesis.

We thank the reviewer for this comment and updated the table and the results accordingly to the suggestion.

1. “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.” Why? Do you mean the group difference on the ASL task?

Thank you for clarifying. Indeed this is what we meant, and we revised the sentence as follows:

Importantly, the ASL-decoding relationship was similarly strong within the DD group alone (R = 0.50, one-tailed p = 0.03, BF10 = 2.04). This indicates that the correlation found in the entire sample was not solely due to the differences in ASL and decoding skills between the TYP and DD groups.

1. Relatedly to the correlational findings, you did not report whether the ASL-decoding correlation survives the correction for mutliple comparisons. Could you please report that? Based on Table 3, it is not significant after the correction. If indeed not, this should be reflected on in the Discussion.

We appreciate the comment. This ASL-decoding correlation across the entire sample survives the correction for multiple comparisons. But within the DD group, due to a small sample size, this relationship did not survive the Bonferonni correction. However, it is important to present the data, because it facilitates our interpretation of the ASL-decoding correlation in the whole sample. We wanted to demonstrate that such a relationship is not purely caused by a bimodal distribution of the two groups on both variables of interest. Instead, the correlation is as strong within the DD group.

1. Why do you report one-tailed p values instead of two-tailed ones?

We had a priori hypotheses regarding the directions of the correlations because better statistical learning could be associated with better reading and language skills but there is no reasonable hypothesis that better statistical learning would be associated with *worse* reading or language skills. We believe it is appropriate and transparent to report the p-values along with the R values. We have now added the a priori hypotheses on page 17 to justify the reporting of one-tailed p-values.

We hypothesized that greater learning performance (more negative slope and higher SL test accuracy) would be associated with better reading and phonological skills.

1. Moreover, on page 17, you report that the ASL-decoding relationship statistics within the DD group are the following: R = 0.50, one-tailed p = 0.03, BF = 2.04. However, if I'm not mistaken, these numbers do not match the statistics in Table 3 (at least the BF).

We thank the reviewer for catching this error. In the process of rerunning the analyses, we caught an error in our R script that altered the number of participants included in the pairwise correlation analyses (due to a file merge error). We fixed the error and updated the statistics in **Tables 2** and **3.** There were no major changes to the main findings as a result of this correction.

1. In the third paragraph of the Discussion, you state that ASL correlated with word decoding, on the whole sample and within the DD group, but not on the TYP group. Then you write that this is consistent with the findings of Qi et al. (2019), where they showed a link in neurotypical adults and children. However, if I understood correctly, in your neurotypical sample, there was no correlation between ASL and word decoding. Hence, the findings are not consistent.

We appreciate the reviewer taking the time to carefully read our previous report on the neurotypical sample. The consistency of our findings across the two studies was discussed at a conceptual level. We used the reading fluency subtest from WJIII to characterize reading skills in our previous study, which was more strongly correlated with ASL than with VSL accuracy. However, the adult sample in the previous study did not have word decoding measures, whereas, in the current study, we did not have the reading fluency subtest measure from WJIII. Nevertheless, our findings are consistent with the previous study in terms of the specific relationship between reading skills and ASL.

We did, however, have overlapping word decoding measures across the current study and the children sample in the previous study. To address Reviewer 1’s comment, we examined the relationship between nonword decoding and ASL RT slope, which was found strongly correlated within children in our previous sample in Qi et al. (2019). As expected, dyslexic adults showed a trend in the same direction. Quicker RT acceleration was marginally related to better nonword decoding (R = -0.40, *p*  = 0.07), but not related to real word decoding measured by Word Identification (r = -0.08, *p*= 0.78). Therefore, we believe the current study replicated our core findings in TYP populations in a different sample. We decided not to include this exploratory analysis into our manuscript due to its marginal nature of the findings but have clarified the consistency of our findings in the discussion.

The statistical learning results further rule out a domain-general procedural learning deficit. We found that  adults with dyslexia showed an impairment in ASL, but intact ability in VSL. Moreover, reading skills, measured by word decoding, were strongly associated with ASL but not with VSL, across the entire sample and in the dyslexia subgroup. These results are consistent with previous research indicating a specific link between ASL and both sentence-level fluency and nonword decoding skills in neurotypical adults and children (Qi et al., 2019). Our study confirmed these findings and further showed that ASL is also linked to word-level decoding skills in adult readers. Interestingly, the relationship between ASL and reading-related skills was even stronger among dyslexic adults than among typical readers in our study. This suggests that the connection between ASL and reading may be particularly relevant in contexts that require more effortful decoding, such as in individuals with dyslexia and younger children, and on nonword reading tasks.

1. Relatedly to the results in point (1), why did you decide on correlate the whole sample and separately the two groups as well?

The correlation analyses across the entire sample examine the learning and reading measures along a continuous spectrum, while correlation analyses within each group revealed important predictors of reading that are more specific to each group.

1. Also relatedy to the ASL-decoding correlation, I'm not really convinced by the correlation in the DD group, the BFs are anecdotal, and as I understand, the correlation does not survive the correction for multiple comparisons. This should be mentioned in the Discussion.

As we explained in our response to Reviewer 1’s comments # 6 and # 7, the ASD-decoding relationship within the DD group supports the relationship between individual differences in ASL and word decoding. We have now toned down the specificity of the SL-decoding relationship for dyslexic participants in the Discussion.

1. On page 20, you explain the intact VSL in DD with the possibility of awareness. Did you measure explicit knowledge of the triplets? I agree that the target-detection task might have boosted attention, but I’m not sure that would also enhance the explicitness of the underlying regularities.

We regretfully did not measure participants’ explicit knowledge of the triplets' patterns. So this would definitely be an interesting future direction.

1. It’s a very minor note, but I would follow the same order when introducing the tasks as the order in the results section.

We thank the reviewer for the comment and updated the order in Methods to match the task order in the Results.

**Reviewer #2**

1. This is an interesting and well-conducted experiment that can potentially add novel knowledge to the literature. However, I have two main concerns. The first one is that the level of difficulty is not equated between the VSL and ASL tests. As a consequence, it is difficult to conclude that statistical learning in the auditory modality is specifically impaired in adults with DD. For this conclusion to hold, it would be necessary to have similar levels of performance at the end of the familiarization phase in the VSL and ASL tests. The second one is that results are not clear-cut. If deficits in ASL, but not in VSL, play a central role in adults with DD, positive correlations should be found between the level of performance in the ASL task (but not in the VSL task) and the level of performance in the various tasks used to assess language proficiency. However, results of correlation analyses revealed that this was not the case and that, by contrast, positive correlations were found in adults with DD between VSL and phonological awareness.... It is also surprising that no significant correlations were found between the different learning tasks. The authors should seriously consider these discrepant findings that are not in line with the hypothesis that ASL is “an early step towards phonological awareness, a pivotal building block of literacy development.”  
   The introduction is to the point, concise while providing relevant information. The hypotheses are clearly stated, based on the literature review.

We appreciate the reviewer’s comments and have strengthened our manuscript to address these main concerns. To respond to the reviewer’s concern about task comparability, we have added the comparison of the task demands during familiarization and test phases within the TYP group into the result section. Our typical readers performed similarly during both phases across the two SL tasks, suggesting that the two tasks are equated on task difficulties.

To compare the task demands during the familiarization phase across the two SL tasks, we tested the interaction between task and trial number within the TYP group. The RT changes during the familiarization phase were similar between ASL and VSL tasks (b = 0.0006, SE = 0.004, t = 0.164, p = 0.87, 𝑅2𝑚 = 0).

To compare the task demands during the test phase across the two SL tasks, we tested the main effect of tasks within the TYP group. The TYP group performed similarly across ASL and VSL tasks (b = -0.18, SE = 0.26, t = -0.69, p = 0.49, 𝑅2𝑚 = 0).

To respond to the reviewer's concern, we have provided evidence showing that word decoding skills wit*hin* DD are specifically related to ASL, despite a relatively modest sample size.

The ASL-decoding relationship was similarly strong 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 simply due to the co-existing group differences on both ASL test accuracy and decoding skills.

In the Discussion, we have provided an explanation for the relationship between VSL RT slope and phonological awareness, as well as the relationship between procedural learning slope and word decoding, which are specific to the DD group:

Individuals with dyslexia often use compensatory strategies for reading, such as relying on explicit strategies (Ullman and Pullman, 2015; Moojen et al., 2020), visuospatial abilities (Goulandris & Snowling, 1991; Snowling & Hulme, 1989; Bacon & Handley, 2014), and contextual cues (Perfetti & Roth, 1981; Stanovich, 1980; Ben-Dror et al., 1991; Bruck, 1990). In the current study, we found a significant positive relationship between visual statistical learning and phonological awareness, as well as between learning on the rotary pursuit task and word decoding, but only in the dyslexia group. These findings suggest that individuals who had better skill learning on the rotary pursuit task and those with superior visual statistical skills were better able to develop their reading-related abilities (although the directionality of these effects cannot be established). These correlations raise the possibility that intact procedural and implicit memory mechanisms are resources for compensatory gains in reading rather than part of the reason for reading difficulty in dyslexia. However, as these effects were specific to certain tasks and not others (such as mirror tracing or auditory statistical learning), further research is needed to fully understand the underlying mechanisms of these compensatory pathways.

Lastly, the lack of a relationship between learning tasks was not surprising. Instead, the pattern is consistent with the view that procedural learning and statistical learning are not unitary constructs. Our findings concur with previous SL studies in neurotypical adults (Erickson et al., 2015; Siegelman et al., 2015) as well as classic neuropsychological studies dissociating basal-ganglia-dependent vs. cerebellum-dependent procedural learning (Gabrieli et al.,1997; Laforce Jr & Doyon, 2001).

1. p.7, l.12-13: “pluralist view of SL (Frost et al., 2019) positing that SL across modalities and domains operates through partially overlapping, but distinct mechanisms.”: pls comment on how this may relate to the various forms of dyslexia

We thank the reviewer for this comment. We updated the paragraph as follows:

The overlap between the tasks represents the domain-general implicit learning mechanisms that were hypothesized to be impaired in dyslexia. Therefore, a direct comparison between auditory and visual SL tasks of similar design is necessary to reconcile whether certain types of SL are indeed more vulnerable than others in dyslexia or whether dyslexia is characterized by cross-domain learning deficits.

1. p.8, l.17 : pls specify whether the authors used the direct and/or reverse forms of the Digit Span subtest

Thank you. We clarified that we used the Forward and Backward Digit Span subtests of the WAIS.

1. It is somewhat surprising that 6 adults were not diagnosed as dyslexics before taking part in the experiment: how did they manage throughout their school studies? Were they less impaired that the 20 adults with an external diagnosis of DD? pls comment.

Thank you for bringing up this important point. We conducted a more in-depth evaluation of participant responses on the Adult Reading History Questionnaire and found that 24 out of 29 participants in the DYS group reported some history related to reading difficulties. We updated the manuscript to reflect these numbers. There are various reasons why someone may not receive a formal diagnosis of dyslexia, including socioeconomic factors (not having the resources to pursue a formal evaluation) and lack of awareness (many students are misidentified in the schools). There were two DYS participants who reported no history of reading difficulties. These participants scored within the range of other DD participants across the measures, as shown in the table below.

IQ- Mean (SD): 108.50 (9.19)

WID- Mean (SD): 99.50 (19.09)

WA- Mean (SD): 93.00 (19.80)

SWE- Mean (SD): 87.00 (1.41)

PDE- Mean (SD): 86.50 (2.12)

Vocabulary- Mean (SD): 102.50 (12.02)

Blending- Mean (SD): 9.50 (2.12)

Nonword- Mean (SD): 5.50 (0.71)

Elision- Mean (SD): 9.00 (1.41)

DigitsSpan- Mean (SD): 12.00 (5.66)

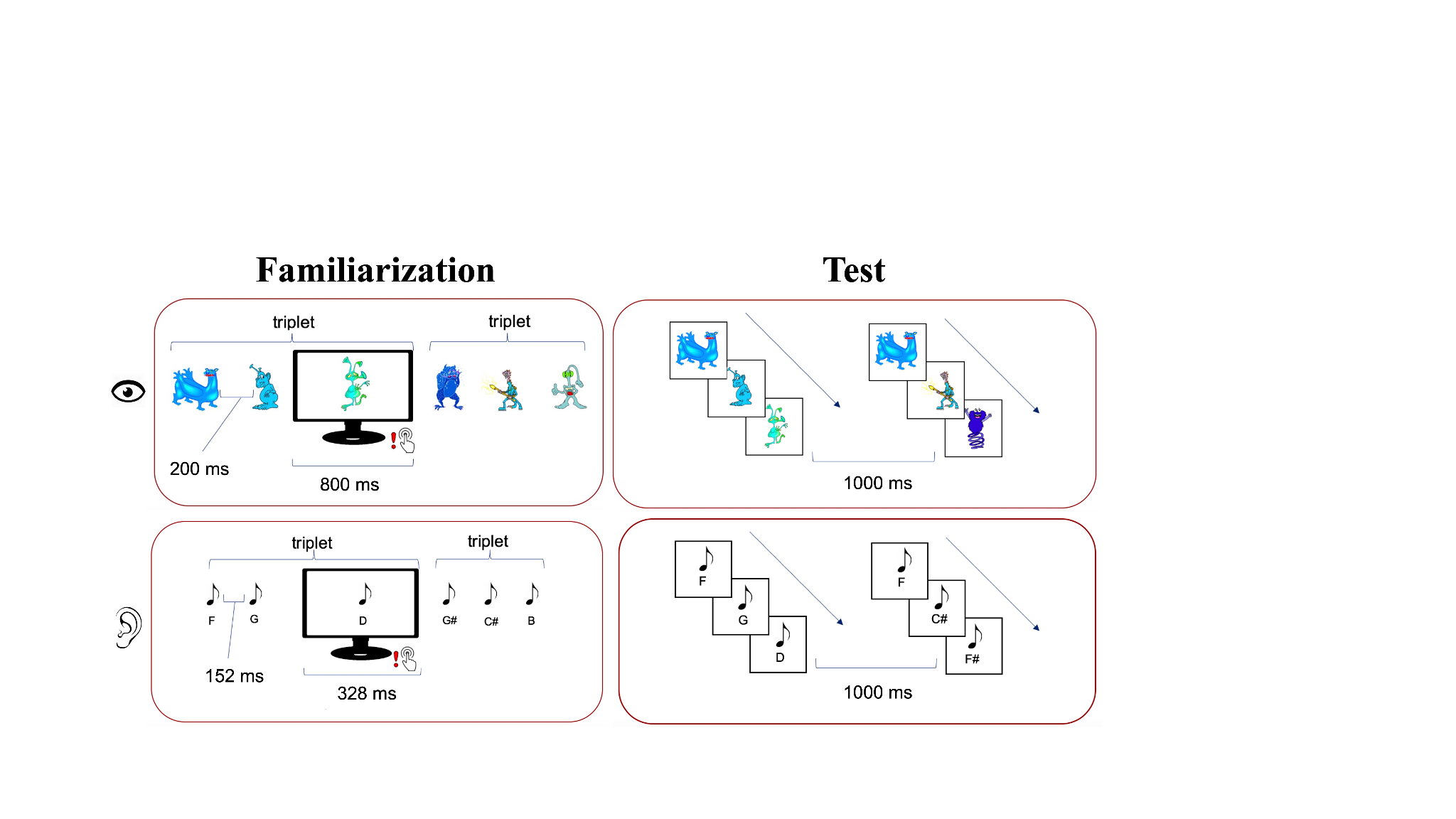
1. p.9: As shown in Table 1, adult with DD performed significantly worse than typical readers on all tested dimensions, except Age and IQ: is it taken to indicate that DD is a general deficit in the various aspects of language and short-term/working memory processing that were tested in this experiment?

Thank you for pointing this out. The observed pattern of results, indicating that adults with DD performed significantly worse than typical readers on reading, language, and phonological measures (elision, blending, and short-term memory), is consistent with previous literature and the theoretical conceptualization of dyslexia as a deficit in phonological and decoding skills (Brady et al., 1983; Shankweiler et al., 1979; Snowling, 2000; Stanovich & Siegel, 1994; Vellutino et al., 1994). Moreover, the measures used in this study were specifically selected based on previous research that has documented deficits in these cognitive processes in individuals with dyslexia. Lower vocabulary scores observed in individuals with DD may reflect reduced reading experience, which is often associated with the disorder. To address this point, we added the following in the Results section:

These findings are consistent with previous literature and the theoretical conceptualization of dyslexia as a deficit in phonological and decoding skills (Brady et al., 1983; Shankweiler et al., 1979; Snowling, 2000; Stanovich & Siegel, 1994; Vellutino et al., 1994). Moreover, the lower vocabulary scores observed in individuals with DD may reflect reduced reading experience, which is often associated with the disorder (Van Der Kleij et al., 2022; Stanovich, 1986).

1. p.9: the description of the different tasks is clear, but it could be nice to illustrate the VSL and ASL tasks on a figure.

We thank the reviewer for this suggestion. Accordingly, we added Figure 1, which illustrates and describes Visual statistical learning and auditory statistical learning task procedure.



​​**Figure 1**. Visual statistical learning (VSL) and auditory statistical learning (ASL) task procedure. During the familiarization phase (left), participants were exposed to a continuous stream of visual (upper) or auditory (lower) stimuli containing triplet patterns while performing a target detection cover task. During the test phase (right), participants completed a two-alternative forced-choice task. Figure adapted from Qi et al. (2019).

1. p.11: could you pls explain in a few words the advantages of Bayesians models?

Thank you. We revised the description of the advantages of Bayes Factors for correlations as follows:

Bayesian models provide good precision even in smaller data sets (Lee & Song, 2004). Importantly, Bayes factors provide a measure of how likely the data are 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. Additionally, Bayesian Factors are less prone to type I errors than traditional hypothesis testing methods, as they provide a continuous measure of evidence for both the null and alternative hypotheses. Based on 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).

1. p.14: it seems that there is a discrepancy between the title of Figure B (err) and C (time) and the information given in ordinate. Pls modify. Pls also present results and figures in the same order as followed in the Methods section (Mirror tracing, Rotary pursuit, VSL and ASL)

We thank the reviewer for this comment. We updated the task description order to match that of the Results section and revised the Figure 1 (now Figure 2) legend as follows:

Performance across trials on procedural learning and implicit statistical learning tasks for adults with dyslexia (DD, solid lines) and typical readers (TYP, dashed lines). (A) Rotary Pursuit proportion on target; (B, C) Mirror Tracing completion time and number of errors; (D) Auditory Statistical Learning familiarization phase response time; and (E) Visual Statistical Learning familiarization phase response time. The vertical dash-dotted lines in 1A-1C indicate task breaks.

1. p.13: “Eight participants (2 in the DD group and 6 in the TYP group) were removed from the analyses for having fewer than 12 (25%) valid responses across 48 target presentations”: this is quite high. Could you pls explain why? And why was the hit rate higher in the VSL task?

We have addressed this comment in our response to a similar comment from Reviewer 1.

As both Reviewer 1 and Reviewer 2 pointed out, this threshold might be too stringent and has caused us to remove too many participants from the slope analyses. Therefore, for group comparison, we followed the Reviewer 1’s suggestion by including all participants into the RT analyses. The results stayed the same. But to avoid skewing our correlational analyses, we opted to remove one TYP participant who showed an unusual deceleration response over the course of exposure (RT slope was more than 3 SD above the sample mean) from the correlations that involve ASL RT slopes. These changes do not affect our main results, because all the ASL 2AFC analyses had already included every participant, regardless of their behavior during the exposure phase.

1. p.13: “Post-hoc within-group analyses” are not justified since the interaction is not significant (p<.087). pls suppress.

We agree that the interaction did not motivate the post-hoc within-group analyses and have deleted these analyses.

1. p.16: “The online learning improvements in response time were marginally larger in VSL than in ASL (three-way interaction between trial number, task, and group: b = 0.02, SE = 0.01, t= 1.87, p = 0.062, 2 = 0.002), suggesting a specific advantage in VSL in DD.”: again, this finding was not significant and should not be discussed.

We have revised this paragraph to improve its clarity (also responding to a similar comment from Reviewer 1):

To compare the degree of group differences in response-time changes across the two SL tasks, we tested the interaction between trial number, task, and group in a linear mixed model. Our analysis revealed a marginal three-way interaction (b = 0.01, SE = 0.005, t = 1.84, p = 0.066, 𝑅2𝑚 = 0.002), that is, the group difference (DD quicker than TYP) in real-time VSL is marginally larger than the group difference in real-time ASL.

1. p19: “Better 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).”: was the 3-way interaction decoding by task (VSL vs ASL) by Group (Typ vs DD) significant?

We thank the reviewer for this comment. Yes, we tested the difference between the two correlations and confirmed that decoding skill was more strongly correlated with ASL than with VSL. We have added this finding to strengthen our manuscript:

1. “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.” I am not sure to understand the logic: could you pls explain?

This comment was also brought up by Reviewer 1. We have now clarified the wording in our result section. Note that we have updated BF values to be one-tailed as well, so that they are consistent with our p-value settings.

Importantly, the ASL-decoding relationship was similarly strong within the DD group alone (0.50, one-tailed p = 0.03, BF10 = 3.28).  This indicates that the correlation found in the entire sample was not solely due to the differences in ASL and decoding skills between the TYP and DD groups.

1. p.21: The statements that “learning deficits in dyslexic adults are specific to the domain that shows the most consistent impairment in dyslexia: auditory processing.” and that “The statistical learning results further rule out a domain-general procedural learning deficit.”, possibly need to be tuned-down if, as mentioned in the general comments, results are linked to the ASL task being more difficult than the VSL task? May be because stimuli in the VSL are more salient than in the ASL tasks? Also, in the ASL task, half of the adults with DD are at chance level: pls comment.

In our response to Reviewer 2’s first general comment, we have provided evidence from the TYP group supporting equated task difficulty across the two SL tasks. Our core finding lies in the fact that the task difficulty is *not* equated in the DD group. As Reviewer 2 points out, many individuals in the DD group performed at chance for the ASL task, despite above-chance performance at the group level. These individuals, who did not seem to learn the ASL triplet patterns, contributed to the group differences in ASL accuracy between TYP and DD. However, we are aware that these findings can be specific to certain languages and samples. We have tuned down our statements as suggested by Reviewer 2.

1. Conclusion: “converging evidence against a domain-general procedural learning deficit in dyslexic adults.” As mentioned above, this conclusion needs to be revised.
2. Table S2: I am surprised that the authors used the term “race” to differentiate types of human beings. It seems to me that “groups”, “ethnic groups” or “populations” are better terms.

Thank you. We now say “racial groups’ in the S2 Table. By NIH criteria, it is common in the US to distinguish between race and ethnicity.

1. For income, pls specify that the income is per year and the unit ($)? For education, pls specify what the percentages refer to?

Thank you we clarified this information in the S2 Table.

**Reviewer #3**

1. How did you decide on sample size? Given the number of correlations calculated, the sample seems small-is (even with corrections for multiple comparisons)

We thank the reviewers for this important question. We updated the Methods section to include the following:

The experiments presented in the current manuscript are part of a larger study investigating behavioral and neural differences in developmental dyslexia. We conducted a power analysis to determine whether the sample size in the current study was sufficient to detect moderate to strong effect size differences between the groups. We hypothesized that implicit learning is a causal factor in dyslexia, and therefore, group differences should be similar in magnitude to those observed for phonological awareness, which is a well-documented deficit in dyslexia. Specifically, the effect size of phonological awareness differences has been reported to be d = -1.37 (Melby-Lervåg, Lyster & Hulme, 2012). Based on our power analysis results, which were based on similar-sized effects, we concluded that a sample size of 25 participants per group would be adequate to detect a large effect size with a power of 0.8 and a significance level of 0.05.

1. For the familiarization phase of the SL tasks – how were the ‘targets’ explained to participants?

We have provided more details in the method section.

In the familiarization phase, participants were shown a target stimulus during the instructions and were told to press the space bar as soon as they saw or heard the same target during the familiarization phase.

1. What does R^2m stand for (e.g., p. 10 line 51 and elsewhere)?

These values represent marginal R squared, the proportion of variance explained by each fixed effect in the linear mixed models. These values were computed via the coefficient of determination (R squared) using the beta function of the package glmm (Jaeger, 2017; Nakagawa & Schielzeth, 2013).

1. Methods: p. 11 - 12. Some participants were omitted from the analysis of the familiarization phase of the ASL task It is not clear whether the same participants were included in the analysis of the test phase data even though their data for the "learning" phase was invalid.

We appreciate the reviewer's comment. Both Reviewer 1 and Reviewer 2 highlighted the issue of removing too many participants from the ASL RT slope analyses due to the hit rate threshold. To address this concern, we followed Reviewer 1's suggestion and included all participants in the RT analyses for group comparison, which did not alter our results. However, to prevent distortion in our correlational analyses, we excluded one TYP participant with an atypical deceleration response during exposure (RT slope more than 3 SD above the sample mean) from correlations involving ASL RT slopes. Importantly, these adjustments do not impact our main findings, as all ASL 2AFC analyses had already incorporated every participant, irrespective of their performance during the exposure phase.

1. P. 20 lines 17-19 the papers cited are not specific to dyslexia.

Thank you. Indeed, some of the papers we cite examined implicit/procedural learning in the context of language-based disorders, including developmental dyslexia more broadly. These papers, however, do make explicit links between deficits in the procedural system and reading impairments in dyslexia.

1. P. 20/ Auditory learning in dyslexia has also been reported to be impaired for other tasks: More specific perceptual learning (Gabay, Karni & Banai, 2018) and category learning (Gertsovski & Ahissar, 2022).

We thank the reviewer for suggesting these papers, and we added them to the Discussion. If we understood the Gabay et al., study correctly, it found that despite initial differences in performance, the dyslexia group showed comparable learning to that of the typical reader group.