Received: 13 August 2019 | Revised: 10 December 2020 | Accepted: 25 January 2021

DOI: 10.1002/dys.1678

RESEARCH ARTICLE

WILEY

The contribution of individual differences in statistical learning to reading and spelling performance in children with and without dyslexia

Merel van Witteloostuijn[1] | Paul Boersma[1] | Frank Wijnen[2] | Judith Rispens[1]

[1]Amsterdam Center for Language and Communication, University of Amsterdam, Amsterdam, The Netherlands

[2]Utrecht Institute of Linguistics OTS, Utrecht University, Utrecht, The Netherlands

Correspondence
Merel van Witteloostuijn, University of Amsterdam, Spuistraat 134, Room 4.17, 1012 VB Amsterdam, The Netherlands.
Email: j.e.rispens@uva.nl

Using an individual differences approach in children with and without dyslexia, this study investigated the hypothesized relationship between statistical learning ability and literacy (reading and spelling) skills. We examined the clinical relevance of statistical learning (serial reaction time and visual statistical learning tasks) by controlling for potential confounds at the participant level (e.g., non-verbal reasoning, attention and phonological skills including rapid automatized naming and phonological short-term memory). A 100 Dutch-speaking 8- to 11-year-old children with and without dyslexia participated (50 per group), see also van Witteloostuijn et al. (2019) for a study with the same participants. No evidence of a relationship between statistical learning and literacy skills is found above and beyond participant-level variables. Suggestions from the literature that the link between statistical learning and literacy attainment, and therefore its clinical relevance, might be small and strongly influenced by methodological differences between studies are not contradicted by our findings.

KEYWORDS
dyslexia, spelling, statistical learning, word reading

1 | INTRODUCTION

Reading and spelling skills are crucial for academic success and large individual differences in literacy attainment exist, with dyslexia affecting around 3–10% of the population (e.g. Miles, 2004; Siegel, 2006). Learning to read involves the mapping from letters or letter clusters (i.e., graphemes) to sounds (i.e., phonemes), while spelling involves the same mapping in the reversed order. Ideally, the correspondences between graphemes and phonemes are one-to-one. In many orthographies, however, this mapping is complex: graphemes can refer to multiple phonemes and vice versa. For example, the grapheme 'c' in English can be expressed either as the phoneme /s/ or /k/ depending on its context (e.g., *cent* vs. *can't*). Although children receive explicit instructions regarding some grapheme-phoneme correspondence patterns in school, their ability to implicitly detect statistical regularities, henceforth 'statistical learning' (SL), has been proposed as an important underlying learning mechanism. This ability is thought to aid the detection of regularities in grapheme-phoneme correspondences when learning to read and spell (e.g., Arciuli, 2017, 2018; Arciuli & Simpson, 2012; Frost, Siegelman, Narkiss, & Afek, 2013; Treiman, 2018). A domain-general learning deficit has been proposed to be the underlying problem in individuals with dyslexia (Nicolson & Fawcett, 2007, 2011); including problems in the area of SL (e.g., Gabay, Thiessen, & Holt, 2015).

One approach to investigating these hypotheses is to correlate performance on independent SL measures with literacy scores. Studies have used a range of SL tasks, including the visual and auditory SL (VSL and ASL, respectively) and serial reaction time (SRT) tasks. Importantly, these tasks all measure participants' ability to implicitly track statistical regularities from the input. Consistent with the above-mentioned proposals, performance on such tasks has been shown to correlate with word and sentence reading in English-speaking adults and children (VSL: Arciuli & Simpson, 2012; ASL: Qi, Sanchez Araujo, Georgan, Gabrieli, & Arciuli, 2019) and with reading Hebrew as a second language in adults (Frost et al., 2013). Replicating Arciuli and Simpson (2012), the relationship between VSL performance and reading accuracy was shown in Norwegian-speaking children (von Koss Torkildsen, Arciuli, & Wie, 2019). Findings by Hung et al. (2018) confirmed this relationship using the SRT task in a group of English-speaking adolescents. A second approach to studying the relationship between SL and literacy attainment is to compare the SL performance of children and adults with dyslexia to typically developing (TD) peers. In line with the hypothesized SL deficit, several studies using a range of SL measures report that individuals with dyslexia perform poorly relative to control groups (e.g., adults: Menghini, Hagberg, Caltagirone, Petrosini, & Vicari, 2006; Sigurdardottir et al., 2017; children: Gabay et al., 2015; Jiménez-Fernández, Vaquero, Jiménez, & Defior, 2011; Singh, Walk, & Conway, 2018).

Not all studies detect this relationship between SL and literacy skills. For example, in a large sample of English-speaking TD children (N = 101), no correlations were observed between SL and measures of reading and spelling (West, Vadillo, Shanks, & Hulme, 2018). In a follow-up study, the authors again found no evidence for a relationship between SL and reading, at least when attention was controlled for (West, Shanks, & Hulme, 2018). Also, Schmalz, Moll, Mulatti, and Schulte-Körne (2018) did not find evidence of the relationship between SL tasks and reading ability in German-speaking adults. On the one hand, large differences in p-values are expected just by chance when recruiting a new sample of participants. On the other hand, Schmalz et al. (2018) suggest that failures to replicate the correlation between SL and reading are possibly due to the use of different measures of SL, since low correlations between such measures have been previously reported (e.g., Capel, 2018; Misyak & Christiansen, 2012; Schmalz et al., 2018; Siegelman & Frost, 2015). Furthermore, these null findings have led to questions regarding the reliability of statistical learning measures (e.g., Siegelman, Bogaerts, Christiansen, & Frost, 2017; West et al., 2018a), especially in use with child participants (Arnon, 2020, 2019b). A study examining the link between learning on the SRT task and a range of language skills in English-speaking children with and without developmental language disorder (DLD) similarly found no correlation between SRT performance and reading words or pseudo-words in either group (Clark & Lum, 2017). Null findings also exist regarding SL performance in dyslexia: a number of studies did not find evidence for a difference in performance between participants with dyslexia and non-impaired controls (adults: e.g., Kelly, Griffiths, & Frith, 2002; Pothos & Kirk, 2004; children: e.g., Nigro, Jiménez-Fernández, Simpson, & Defior, 2016; Staels & Van den Broeck, 2017), even when using a range of different SL measures within the same pool of participants (adults: Rüsseler, Gerth, & Münte, 2006; children: van Witteloostuijn et al., 2019). This mix of high and low p-values (i.e., some studies reporting significant group effects and other studies finding null results), though expected for effects with

average detectability, has resulted in literature reviews and meta-analyses in the area of SL in dyslexia to determine the overall effect size (Lum, Ullman, & Conti-Ramsden, 2013; Schmalz, Altoè, & Mulatti, 2017; van Witteloostuijn, Boersma, Wijnen, & Rispens, 2017). Although findings from these meta-analyses suggest that individuals with dyslexia may experience problems with SL when collapsing overstudies, authors have raised the issue of a publication bias in the field that likely inflates the observed effect size in meta-analyses (Schmalz et al., 2017; van Witteloostuijn et al., 2017).

Recently, authors have also stressed the need for research combining the two approaches—correlational studies and the investigation of SL in dyslexia—to further elucidate the relationship between SL and literacy skills (Arciuli, 2018; Arciuli & Conway, 2018). To date, several studies have indicated that SL performance relates to reading ability in participants with and without dyslexia (English adults: Gabay et al., 2015; Howard Jr, Howard, Japikse, & Eden, 2006; Icelandic adults: Sigurdardottir et al., 2017; Hebrew children: Vakil, Lowe, & Goldfus, 2015; Swedish children: Hedenius et al., 2013; Dutch children: van der Kleij, Groen, Segers, & Verhoeven, 2019). The fact that replication with a sample of Spanish-speaking children (Nigro et al., 2016) led to a null result, does not contradict the statistically significant findings above, but Nigro et al. (2016) argue, based on this *p*-value difference, that SL may play a less prominent role in learning to read a shallow orthography such as Spanish (i.e., a writing system with a relatively transparent grapheme-to-phoneme mapping; although see conflicting results in other more transparent orthographies such as Icelandic, Swedish and Dutch by Sigurdardottir et al., 2017, Hedenius et al., 2017, and van der Kleij et al., 2019, respectively).

To summarize: although theory suggests a link between SL and literacy skills, the evidence is weak given the differences between *p*-values. While such differences are expected in repeated sampling, the literature has proposed actual causes for these differences, including differences in SL tasks used (Schmalz et al., 2018), potential confounds at the participant level (e.g., attention, West et al., 2018b) and low reliability of statistical learning measures (e.g., Arnon, 2020, 2019b; Siegelman, Bogaerts, & Frost, 2017; West et al., 2018a). Interestingly, the majority of studies investigating the relationship between SL and literacy skills have done so through simple correlations, not considering participant-level variables (i.e., potential confounds) or other known predictors of reading see Qi et al., 2019; Von Koss Torkildsen et al., 2019). Furthermore, studies of individual differences in statistical learning ability in relation to literacy skills have often not reported the reliability of the statistical learning measures used (e.g., West et al., 2018a; 2018b). Moreover, studies to date have largely focused on reading, thereby disregarding spelling despite its theorized link with SL (Treiman, 2018) and despite the spelling difficulties associated with dyslexia (American Psychiatric Association, 2013). In relation to dyslexia, it is important to elucidate the clinical relevance of SL in literacy acquisition (see for example, Plante & Gómez, 2018, for a discussion regarding the clinical relevance of SL for treatment applications with DLD), since suggestions have been voiced that the true correlation may only be small (or maybe largely mediated by confounding variables).

1.1 | The present study

In the present study, we aim to further investigate the relationship between SL and literacy abilities in children with and without dyslexia. We hope to do so comprehensively and reliably by looking at both reading and spelling performance and using two SL measures that have previously been linked to individual differences in literacy skills: the SRT task (e.g., Hedenius et al., 2013; Howard Jr et al., 2006; Van der Kleij, Groen, Segers, & Verhoeven, 2018) and VSL task (e.g., Arciuli & Simpson, 2012; von Koss Torkildsen et al., 2019). Additionally, we aim to account for participant-level characteristics (age, gender, socio-economic status and diagnosis), general cognitive skills (nonverbal reasoning and sustained attention) and other known predictors of literacy outcomes such as rapid automatized naming (RAN), phonological processing and phonological short-term memory (e.g., de Bree, Wijnen, & Gerrits, 2010; de Jong & van der Leij, 1999; Furnes & Samuelsson, 2010; Swanson & Howell, 2001; Van Setten et al., 2017; see also Snowling & Melby-Lervåg, 2016 for a meta-analysis). Finally, we report the split-half reliability of the statistical learning measures used in the present study as an indication of their internal consistency and reliability (e.g., Arnon, 2020; Siegelman, Bogaerts, Christiansen, & Frost, 2017). Since the statistical learning measures used in the present study were adapted for use with child participants, we hope to find split-half reliability coefficients that approach the psychometric standard of around $r = 0.80$ (Nunnally & Bernstein, 2004; Streiner, 2003).

Through a regression analysis, we examined the contributions of SL and the aforementioned predictors to reading and spelling performance in 100 Dutch-speaking school-aged children with and without dyslexia. The current study analyses the performance of children who were also reported on by van Witteloostuijn et al. (2019). This previous publication focused exclusively on group differences (developmental dyslexia vs. children without dyslexia) in SL abilities. In contrast, in the current study, we focus on the contribution of individual differences in SL and other cognitive factors to variability in literacy attainment. The research questions were as follows:

1. Does SL ability (SRT and VSL tasks) contribute to literacy performance, taking into account non-linguistic cognitive (non-verbal reasoning and attention) and phonological skills[1] (RAN letters and pictures, non-word repetition and digit span forward tasks)?
   And, if so,

2. Are the contributions of phonological skills and/or SL different for
   a. children with and without dyslexia?
   b. reading and spelling?

2 | METHODS

2.1 | Participants

Fifty children with a diagnosis of dyslexia (26 girls, 24 boys, age range 8;4–11;2, *M* = 9;10) and 50 age-matched control children (24 girls, 26 boys, age range 8;3–11;2, *M* = 9;8) in grades three to five participated. Ten additional children with dyslexia and four additional control children were tested but turned out not to meet our pre-determined inclusion criteria (dyslexia: norm score of at most 6 (i.e., 10th percentile) on word and pseudo-word reading; control group: norm score of at least 8 (i.e., 25th percentile)). All 100 children that participated in the present study completed each of the tasks as outlined in Section 2.2. Children with dyslexia were recruited through treatment centers and Facebook support groups for parents, while children in the control group were recruited through primary schools. All parents and children consented to participation prior to testing in accordance with the ethics review board of the (name deleted to maintain the integrity of the review process). All participants were native speakers of Dutch and none had been diagnosed with (additional) developmental disorders as reported by parents (in the case of participants with dyslexia) and teachers (in the case of control participants). Note that the sample reported here is identical to the sample reported by van Witteloostuijn et al. (2019), since, as previously stated, the current study is a re-analysis of the same sample. Also, the control group partly overlaps with studies investigating SL and its relationship with language in children with DLD (Lammertink et al., 2020; Lammertink et al., 2020).

Table 1 presents descriptive statistics of participant characteristics. The children with and without dyslexia were found not to differ significantly from one another regarding their age (t = 0.839, p = .40), socio-economic status (t = 0.173, p = .86) and non-verbal reasoning skills (t = 0.041, p = .97). The children with dyslexia achieved marginally significantly lower scores than their TD peers on sustained attention (t = 1.939, p = .055). These participant characteristics are included in our regression analyses as control variables.

2.2 | Materials

2.2.1 | Literacy measures

Children's technical reading skills were assessed using two standardized tests: the reading of single Dutch real words (Brus & Voeten, 1972) and pseudo-words (Van den Bos, Spelberg, Scheepsma, & Vries, 1994). The task was to read

TABLE 1 Descriptive statistics of background measures for children with and without dyslexia

	Dyslexia (*N* = 50)			
	Raw	**Standardized**	**Raw**	**Standardized**
Female: male	26:24		24:26	
Age	9;10 (0;9)	N/A	9;8 (0;10)	N/A
SES[a]	0.2 (1.2)	N/A	0.2 (1.1)	N/A
Nonverbal reasoning[b]	37.2 (6.6)	55.7 (25.0)	37.3 (8.1)	60.1 (28.1)
Sustained attention[c]	7.0 (2.5)	7.4 (3.3)	7.8 (1.8)	9.1 (3.0)

[a]SES was determined on the basis of postal codes through the *Netherlands Institute for Social Research*.
[b]Non-verbal reasoning was assessed through *Raven's Standard Progressive Matrices* (Raven & Raven, 2003). Raw and standardized scores on nonverbal reasoning represent the number of items answered correctly out of 60 and percentile scores (norm = 50), respectively.
[c]Sustained attention was measured using the *Score!* subtest of the Dutch *Test of Everyday Attention for Children* (Schittekatte, Groenvynck, Fontaine & Dekker 2007). Raw and standardized scores on sustained attention represent the number of items answered correctly out of 10 and norm scores (norm = 10), respectively.

(pseudo-)words as quickly and accurately as possible from a set list of words in a set amount of time (1 min for words, 2 min for pseudo-words), and thus the outcome measure was the number of (pseudo-)words read correctly within the time limit. Dutch spelling proficiency was measured through a standardized dictation test consisting of six blocks of 15 words each (not including verbs) of increasing difficulty both between and within blocks (Braams & Vos, 2015). Each child completed two blocks depending on their grade in school (i.e., the spelling score is the number of words spelled correctly out of 30). Every word was presented orally in a context sentence after which the target word was repeated and the child was required to manually write down the word according to Dutch spelling.

2.2.2 | Phonological skills

First, children were tested on two subtests of RAN: one containing letters and one containing pictures of common objects (standardized test; see Van den Bos & Spelberg, 2007). Children were instructed to name the letters or pictures as quickly and accurately as possible. Next, children's phonological processing and short-term memory were assessed through two tasks: a short (not norm-referenced) non-word repetition task (NWR-S: le Clercq et al., 2017) and a standardized forward digit span task (Kort, Schittekatte, & Compaan, 2008). In the NWR-S, children listened to 22 pre-recorded non-words and had to repeat them as accurately as possible. All non-words were between three and five syllables long and were either phonologically likely or unlikely according to Dutch phonotactic probabilities. Children's responses were recorded and scored as either correct or incorrect. In the forward digit span task, children had to repeat sequences of digits of increasing length (2–9 digits) in the correct order. Each level of the task contained two items; to advance to the next level, the child had to answer at least one out of two items correctly. Testing was halted once a child answered both items within one level incorrectly.

2.2.3 | Statistical learning

Serial reaction time task

The SRT task used in the present study is identical to the one described by van Witteloostuijn et al. (2019). Participants were exposed to a single visual stimulus that repeatedly appeared in one of four locations (quadrants) on a tablet screen with a 250 milliseconds interval (Nissen & Bullemer, 1987). They were required to respond to the stimulus'

location on the screen by pressing one of four corresponding buttons on a gamepad controller as quickly and as accurately as possible. Without the participants' knowledge, the SRT task was divided into seven blocks. In blocks 2 through 5 and block 7, the stimulus followed a predetermined sequence of 10 locations (4, 2, 3, 1, 2, 4, 3, 1, 4, 3) which was repeated six times in each block (i.e., 60 trials per block), while the stimulus was presented in random order during 60 trials in the intervening block 6. Block 1 consisted of 20 random trials to accustom participants to the task and is not included in the analysis. Learning of the statistical structure in the SRT task is evidenced by longer RTs to random stimuli (block 6) than to structured stimuli in the surrounding sequence blocks (blocks 5 and 7). Individual scores on the SRT task were computed by subtracting the mean normalized RT to structured input (average of blocks 5 and 7) from the mean normalized RT to unstructured input in block 6.

Visual statistical learning task

The VSL task used in the present study is identical to the one described by van Witteloostuijn et al. (2019) and was similar in structure and design to previous studies by Arciuli and Simpson (2011, 2012). Twelve visual stimuli (aliens) were presented one by one on a tablet with touch screen. Unbeknownst to the participants, these 12 stimuli repeatedly appeared in the same four groups of three (i.e., triplets; *ABC*, *DEF*, *GHI* and *JKL*). Learning of this triplet structure was originally assessed through three measures: an online reaction time (RT) and two offline accuracy measures. Since no evidence of learning was found through the online RT measure in children with and without dyslexia (vvan Witteloostuijn et al., 2019), we focus on the offline measures of learning.

Prior to the experiment, children were informed that aliens stood in line to go home with a space ship and that they would see all of the aliens one by one. They were instructed to pay attention to the aliens and were told that some of the aliens liked one another and stood in line together. The exposure phase of the VSL task contained four separate blocks, each consisting of six occurrences of each triplet. The same triplet never appeared twice in a row and triplet pairs were never repeated (Arciuli & Simpson, 2011, 2012; Turk-Browne, Jungé, & Scholl, 2005). In between blocks, participants received stickers on a diploma. A cover task was inserted in the exposure phase to ensure that children paid attention to the stimulus stream (Arciuli & Simpson, 2011, 2012). Three individual stimuli per block appeared twice in a row and children had to respond to a repeated stimulus by pressing the alien on the screen. Each stimulus within each triplet was repeated once during the exposure phase (e.g., the triplet *ABC* occurs once as *AABC*, *ABBC* and *ABCC*) and three distinct triplets contained a repetition in random positions in each of the four blocks, again all three stimulus positions within triplets once (e.g., *AABC*, *DEEF*and *GHII*).

Subsequent to exposure, children were tested on their knowledge of the triplet structure. Using the same set of 12 visual stimuli, four foil triplets (*AEI*, *DHL*, *GKC* and *JBF*) were created for use in the offline test phase that consisted of 40 multiple-choice questions. A 16 three-alternative forced-choice (3-AFC) questions in which children had to fill in a missing stimulus (chance level = 33.3%) were followed by 24 two-alternative forced-choice (2-AFC) questions in which they had to pick the more familiar group of aliens (chance level = 50%). Both 3-AFC and 2-AFC question blocks were introduced through two practice items during which children were encouraged to make a guess when they were uncertain of the correct response. Individual scores on the VSL task represent the number of items answered correctly out of 16 and 24 on 3-AFC and 2-AFC questions, respectively.

2.3 | General procedure

The SRT and VSL tasks were programmed and run using E-prime 2.0 (Psychology Software Tools, 2012; Schneider, Eschman, & Zuccolotto, 2012) on a Windows Surface 3 tablet with touch screen. Pre-recorded auditory instructions (SRT) and stimuli (NWR-S) were played over Sennheiser HD 201 headphones. Responses in the SRT tasks were given through a Trust wired GXT540 gamepad controller. Children's responses during the reading, RAN and NWR-S tasks were recorded using an Olympus DP-211 voice recorder.

As previously mentioned in Section 2.1, children were tested in the context of a larger project. An experimenter administered a battery of tasks one-on-one in a quiet room either at the child's home or school. Testing took place in three sessions that each lasted around an hour. The SL tasks were tested in separate sessions along with a number of other measures. The order of the test sessions and the tasks within sessions were counter-balanced: participants were randomly assigned to one out of six testing orders.

2.4 | Data scoring and analysis

We performed a linear regression analysis through the *lm* function in *R* software to assess the contribution of a number of predictors in explaining individual variation in reading and spelling attainment combined in a single model. Confidence intervals (CIs, 95%) were computed by the profiling method (through the *confint* function) and were used to compare the contribution of predictors to reading versus spelling (research question 3b). Predictors in the model included control variables (group membership, age, gender, SES, non-verbal reasoning and sustained attention), phonological skill measures (RAN letters, RAN pictures, NWR-S and digit span forward) and measures of SL (SRT and VSL). Interactions between group and phonological skills and between group and SL measures were investigated (research question 3a). Significance of individual predictors to reading and spelling combined was assessed through the *MANOVA* function in the *car* package (version 2.1–5; Fox et al., 2012). To answer the first two research questions, we conducted model comparisons between the full model and models from which (a) phonological skill measures and (b) SL measures were removed.

All raw scores on continuous measures were centered and scaled using the *scale* function. Categorical predictors were coded into orthogonal contrasts: gender was coded such that females were marked as −1/2 and males were marked as +1/2; group membership was coded such that the control group was marked as −1/2 and the dyslexia group was marked as +1/2. Finally, since both reading and VSL were measured through two subtests (reading words and pseudo-words; VSL 3-AFC and 2-AFC), the averages of the centered and scaled subtests were used in our analyses. Summary level data and *R* Markdown and html files detailing our analyses are available on the Open Science Framework (https://osf.io/dr72a/).[2]

The split-half reliability of the statistical learning tasks was computed using Spearman–Brown corrected Pearson correlations (see also Arnon, 2020; Siegelman, Bogaerts, & Frost, 2017). In the SRT task, the split-half reliability was calculated for each individual as the correlation between the difference in RT between the random stimuli (block 6) and structured stimuli in the surrounding sequence blocks (i.e., blocks 5 and 7) in *even* versus *odd* trials. This difference in RT was obtained from the linear mixed-effects model through the random slopes of the relevant predictor (i.e., the difference in RT between random and sequence). Similarly, the correlation between the accuracy on even and odd trials in the VSL offline test phases (2-AFC and 3-AFC) was used to calculate the split-half reliabilities (i.e., the random slopes of the intercept). We would like to refer the reader to our OSF project page for more detail regarding the calculations of the split-half reliabilities.

3 | RESULTS

We first provide the descriptive statistics and group comparisons of the outcome measures and predictors included in our linear regression analysis in Section 3.1. The confirmatory analyses aimed at answering our research questions are presented in Section 3.2. These consist of the linear regression analyses and model comparisons. Section 3.3 presents exploratory analyses and findings, which do not provide answers to our research questions but may nonetheless be of interest (cf. Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012). Finally, the results regarding the split-half reliabilities of our statistical learning measures are provided in Section 3.4.

3.1 | Descriptive statistics and group comparisons

Table 2 contains the descriptive statistics of reading, spelling, phonological skills and SL. As expected, our children with dyslexia performed significantly worse than children in our control group on reading (words: t = 13.83, $p = 9.1 \times 10^{-25}$; pseudo-words: t = 16.75, $p = 1.6 \times 10^{-30}$), spelling (t = 11.05, $p = 9.4 \times 10^{-20}$) and phonological skills (RAN letters: t = 5.421, $p = 4.3 \times 10^{-7}$; RAN pictures: t = 4.985, $p = 2.7 \times 10^{-6}$; NWR-S: t = 3.962, p = .00014; digit span forward: t = 5.36, $p = 5.5 \times 10^{-7}$). On average our children learned the statistical structures in the SRT and VSL tasks, and no evidence of a difference in performance between children with and without dyslexia was found on the SL measures (SRT: Δz = −0.027, p = .61; VSL 3-AFC odds ratio estimate = 1.001, p = .996; VSL 2-AFC: odds ratio estimate = 1.076, p = .68; see van Witteloostuijn et al., 2019). For more detail on the analysis of the SL measures, please see the Open Science Framework (https://doi.org/10.17605/OSF.IO/T8SCV).

3.2 | Regression model

The correlations between the measures are included in Appendix A (Table A1, A2). The outcomes of the full model are presented in Table 3 for reading and spelling separately. We now turn to the focus of our study: investigating the contribution of SL to individual differences in literacy attainment, taking into account phonological kills (research question 1) and potential differences between children with and without dyslexia (research question 2a). We see that the main effects of SRT and VSL are non-significant overall (Wilk's λ = 0.97, F(2.80) = 1.164, p = .32 and Wilk's λ = 0.99, F(2.80) = 0.496, p = .61, respectively). The interaction between group and SRT performance was not significant (Wilk's λ = 0.94, F(2.80) = 2.330, p = .10). Although the interaction between group and SRT performance is significant for spelling (*estimate* = 0.29, 95% CI [0.02 … 0.57], p = .036) but not for reading (*estimate* = 0.13, 95% CI [−0.03 … 0.29], p = .11), we cannot infer a difference between reading and spelling due to overlapping 95% CIs

TABLE 2 Descriptive statistics on outcome measures, phonological skills and statistical learning: raw and standardized scores per group

	Dyslexia (N = 50)		Control (N = 50)	
	Raw	Standardized	Raw	Standardized
Reading words[a]	34.1 (11.7)	3.3 (2.1)	66.3 (11.6)	10.5 (2.2)
Reading pseudo-words[a]	22.0 (8.0)	4.4 (1.6)	61.0 (14.4)	11.1 (2.2)
Spelling[b]	8.4 (4.6)	11.8 (13.7)	18.6 (4.7)	49.9 (24.7)
RAN letters[a]	36.1 (10.4)	5.4 (2.7)	27.2 (5.5)	9.6 (3.1)
RAN pictures[a]	53.2 (10.2)	7.7 (2.7)	44.1 (7.3)	10.7 (2.8)
NWR-S[c]	7.3 (2.7)	N/A	9.7 (3.3)	N/A
Digit span forward[a]	7.3 (1.5)	7.7 (2.6)	8.9 (1.5)	10.7 (2.9)
SRT[c]	0.29 (0.28)	N/A	0.27 (0.27)	N/A
VSL 3-AFC[c,d]	8.2 (3.1)	N/A	8.2 (3.8)	N/A
VSL 2-AFC[c,d]	15.3 (4.4)	N/A	15.0 (4.5)	N/A

Note: Raw scores: reading words and pseudo-words = the number of words read within the time limit. Spelling = the number of words spelled correctly out of 30, RAN = the number of seconds spent on the task (i.e., higher score = lower performance). NWR = the number of non-words repeated correctly out of 22. Digit span forward = the number of items answered correctly out of 16. SRT = difference in normalized RTs (RT random − RT sequence). VSL = number of items answered correctly out of 16 (3-AFC) and 24 (2-AFC). Standardized scores represent either [a]norm scores (norm = 10) or [b]percentile scores (norm = 50). [c]No standardized scores are present for the NWR-S, SRT and VSL tasks. [d]Chance level on VSL 3-AFC = 33.3% (5.3 items correct out of 16); 2-AFC 50% (12 items correct out of 24).

(research question 2b). Comparing the full model to a reduced model (where SRT and VSL are removed) does not show a significant effect of the removal of SL measures on the model fit (F[8, 162] = 1.134, p = .34).

Since we specifically wanted to investigate the contribution of SL controlling for potential effects of phonological skills on literacy skills, we will also discuss briefly those variables. From the *MANOVA* results, as presented in Table 4, we see that only RAN letters is a significant contributor to literacy outcomes combined, over and above the other predictors in the model (Wilk's λ = 0.74, F(2.80) = 13.840, $p = 6.9 \times 10^{-6}$). Additionally, the interaction with the group is significant (Wilk's λ = 0.90, F(2.80) = 4.600, p = .013), such that the effect of RAN letters is larger for TD children than for children with dyslexia.[3] The interaction between RAN pictures and group is significant in the same direction (Wilk's λ = 0.92, F(2.80) = 3.453, p = .036). Regarding differences between reading and spelling (research question 2b), the

TABLE 3 Full linear regression model: reading and spelling outcomes separately (*lm*)

	Reading				Spelling			
	b	**95% CI**	***t*-value**	***p***	***b***	**95% CI**	***t*-value**	***p***
Control								
Age	0.092	[0.006 … 0.18]	2.13	.036*	0.014	[−0.14 … 0.16]	0.18	.86
Gender	−0.0011	[−0.17 … 0.17]	−0.012	.99	−0.078	[−0.37 … 0.22]	−0.53	.60
SES	0.018	[−0.07 … 0.10]	0.44	.66	0.018	[−0.13 … 0.16]	0.25	.80
Raven	0.017	[−0.09 … 0.12]	0.32	.75	0.20	[0.03 … 0.38]	2.31	.024*
Attention	0.034	[−0.05 … 0.12]	0.80	.42	−0.020	[−0.16 … 0.13]	−0.27	.79
Group	−1.23	[−1.45 … −1.01]	−11.13	<.001*	−1.48	[−1.86 … −1.10]	−7.77	<.001*
Phonology								
RAN letters	−0.28	[−0.42 … −0.15]	−4.21	<.001*	0.096	[−0.13 … 0.33]	0.83	.41
RAN pictures	−0.10	[−0.21 … 0.007]	−1.86	.066†	0.016	[−0.17 … 0.21]	0.17	.87
NWR-S	0.097	[−0.004 … 0.20]	1.91	.060†	0.072	[−0.10 … 0.25]	0.82	.41
DSF	0.0052	[−0.10 … 0.11]	0.10	.92	0.078	[−0.10 … 0.25]	0.88	.38
SL								
SRT	0.028	[−0.06 … 0.11]	0.67	.51	0.095	[−0.05 … 0.24]	1.32	.19
VSL	0.031	[−0.07 … 0.13]	0.63	.53	−0.017	[−0.18 … 0.15]	−0.21	.84
Interactions								
Group*RAN let	0.27	[0.002 … 0.54]	2.00	.048*	−0.14	[−0.61 … 0.33]	−0.59	.55
Group*RAN pic	0.20	[−0.01 … 0.42]	1.86	.066†	−0.064	[−0.44 … 0.31]	−0.34	.73
Group*NWR-S	−0.067	[−0.26 … 0.13]	−0.67	.50	−0.13	[−0.46 … 0.21]	−0.73	.47
Group*DSF	0.10	[−0.10 … 0.30]	0.99	.33	−0.025	[−0.37 … 0.32]	−0.14	.89
Group*SRT	0.13	[−0.03 … 0.29]	1.64	.11	0.29	[0.02 … 0.57]	2.13	.036*
Group*VSL	0.074	[−0.12 … 0.26]	0.77	.44	0.18	[−0.15 … 0.51]	1.09	.28

Note: RAN let = RAN letters, RAN pic = RAN pictures, DSF = Digit Span Forward. Significant findings ($p \leq .05$) are indicated using an asterisk (*), near-significance ($.05 \leq p \leq .10$) is indicated using a cross (†).

TABLE 4 Full linear regression model: outcomes for reading and spelling combined (*MANOVA*)

	Pillai's trace	*F*(2, 80)	*p*
Control			
Age	0.925	3.24	.044*
Gender	0.995	0.21	.81
SES	0.998	0.09	.91
Raven	0.918	3.57	.033*
Attention	0.981	0.78	.46
Group	0.320	84.88	< .001*
Phonology			
RAN letters	0.743	13.84	< .001*
RAN pictures	0.968	1.30	.28
NWR-S	0.949	2.16	.12
DSF	0.987	0.52	.60
SL			
SRT	0.971	1.16	.32
VSL	0.988	0.50	.61
Interactions			
Group*RAN let	0.897	4.60	.013*
Group*RAN pic	0.921	3.45	.036*
Group*NWR-S	0.992	0.31	.74
Group*DSF	0.977	0.93	.40
Group*SRT	0.945	2.33	.10†
Group*VSL	0.985	0.60	.55

Note: RAN let = RAN letters, RAN pic = RAN pictures, DSF = Digit Span Forward. Significant findings ($p \leq .05$) are indicated using an asterisk (*), near-significance ($.05 \leq p \leq .10$) is indicated using a cross (†).

effect of RAN letters is significantly larger on reading than on spelling since the 95% CIs do not overlap (*estimate* = –0.28, 95% CI [–0.42 … –0.15], $p = 6.5 \times 10^{-5}$ and *estimate* = 0.096, 95% CI [–0.13 … 0.33], $p = .41$, respectively). Importantly, when we compare the full model to a reduced model where the phonological skill measures are removed (RAN letters, RAN pictures, NWR-S and digit span forward), we find that this removal results in a significant decrease in model fit ($F[16,162] = 3.771$, $p = 6.4 \times 10^{-6}$). Thus, taken together, the phonological skill measures used in the present study (RAN letters and pictures, NWR-S and digit span forward) contribute to children's literacy performance.

In sum, there is no evidence that the SRT and VSL measures together contribute to explaining variance in literacy performance in children with and without dyslexia (above and beyond our control variables and phonological skill measures).

3.3 | Exploratory results

3.3.1 | Control variables

As expected, group membership was a significant predictor for both literacy measures combined (Wilk's $\lambda = 0.32$, $F(2.80) = 84.876$, $p = 3.4 \times 10^{-20}$), such that children with dyslexia achieve lower scores than their TD peers. Similarly, age is found to be a significant predictor of literacy performance (Wilk's $\lambda = 0.93$, $F(2.80) = 3.237$, $p = .044$). This effect

is driven by a significant effect of age on reading (*estimate* = 0.092, 95% CI [0.01 … 0.18], t = 2.130, p = .036) but not spelling (*estimate* = 0.014, 95% CI [−0.14 … 0.16], t = 0.181 p = .86). This is to be expected, since the spelling test used is adapted to children's grade, whereas the reading test is not. The opposite pattern is observed for non-verbal reasoning, which is a significant predictor for spelling (*estimate* = 0.20, 95% CI [0.03 … 0.38], t = 2.308, p = .024) but not reading (*estimate* = 0.017, 95% CI [−0.09 … 0.12], t = 0.324, p = .75). Again, the overall effect of non-verbal reasoning on literacy skills combined is found to be significant (Wilk's λ = 0.92, F(2.80) = 3.566, p = .033).

3.3.2 | Phonological skills

In the full model, significant effects are found only for the RAN letters subtest. Therefore, an exploratory analysis was performed to see whether removing the other measures of phonological skills (i.e., RAN pictures, NWR-S and digit span forward) results in a decrease in fit of the model. Results reveal that this is not the case, as the model comparison is not significant (F[12, 162] = 1.559, p = .11). This means that there is no evidence that, taken together, RAN pictures, NWR-S and digit span forward contribute to literacy performance above and beyond RAN letters.

3.3.3 | Statistical learning

Perhaps unexpectedly, we find no evidence that children's VSL performance contributes to literacy scores above and beyond the SRT. To investigate whether the VSL may be of value when the SRT is not considered, we ran an identical model with the SRT measure removed. However, the effects of the VSL remain non-significant both in reading (*estimate* = 0.027, t = 0.555, p = .58) and in spelling (*estimate* = −0.024, t = −0.279, p = .78) and no interactions between VSL and group are found for either outcome measure (*estimate* = 0.077, t = 0.802, p = .43 and *estimate* = 0.19, t = 1.099, p = .27, respectively).

We also wanted to explore the interaction between group and SRT, which approached significance for reading and spelling combined. For further investigation, we performed Pearson's correlations between SRT and our literacy outcomes (see *R* markdown and html files for plots). The correlation with spelling was found to be non-significant in the control group (r = −0.103, p = .48), whereas it reached significance in the group of children with dyslexia (r = 0.372, p = .0078). Similar results are observed regarding reading (control group: r = −0.229, p = .11, dyslexia group: r = 0.348, p = .013).

Finally, as requested by an anonymous reviewer, we investigated the individual scores on the VSL task (even though the results on the VSL showed above-chance performance at the group level). In Appendix B (Figure B1, Table B1), a histogram is included for the 2-AFC and 3-AFC tasks (split out over the two groups) which shows the distribution of the number of items scored correctly. The results reveal that, in the control group, 40 and 44% of participants performed above chance at the individual level on 2-AFC and 3-AFC questions, respectively, while 46 and 50% of participants in the group of children with dyslexia scored above chance. Whether an individual child performed significantly above chance level was determined using the binomial distribution to calculate the number of items that should be answered correctly to reach a p-value smaller than .05 (Siegelman, Bogaerts, & Frost, 2017). For the 2-AFC questions, the individual above-chance performance was reached when 16 (out of 24) or more items were answered correctly; for the 3-AFC questions, this was reached when 9 out of 16 items were answered correctly.

3.4 | Split-half reliability of the statistical learning measures

As explained in Section 2.4, split-half reliabilities were calculated as a measure of the internal consistency and reliability of the statistical learning measures used in the present study. The split-half reliability for the online

measure of learning in the SRT task was found to be $r = 0.71$, 95% CI [0.58, 0.81]. For the offline measures of learning in the VSL task, the split-half reliabilities were $r = 0.70$, 95% CI [0.55 … 0.80] and $r = 0.78$, 95% CI [0.67 … 0.85] for 2-AFC and 3-AFC questions, respectively. Thus, the split-half reliabilities found for the SRT and VSL tasks used in the present study approach the psychometric standard of $r = 0.80$ (see for example, Nunnally & Bernstein, 1994; Streiner, 2003).

4 | DISCUSSION

The current study examined the contribution of SL ability to individual differences in reading and spelling performance in children with and without dyslexia. We aimed to do so while controlling for potential participant level confounds including a range of cognitive and phonological skills, to investigate whether SL contributes to reading and spelling above and beyond other potential predictors of literacy performance. Our finding that phonological skill measures contribute to literacy scores replicates earlier work (e.g., de Jong & van der Leij, 1999; Snowling & Melby-Lervåg, 2016; Swanson & Howell, 2001). Regarding the relationship with SL, exploratory simple correlations suggest that whereas there is a (weak) association between SRT performance and literacy skills in the group of participants with dyslexia, no support for such a link is observed in the control group. No evidence for an association of VSL performance to literacy skills was obtained in either group. However, after controlling for the aforementioned participant-level variables, we find no evidence that SL (SRT and VSL) ability contributes to reading and spelling.[4] Regression analysis did not reveal significant differences regarding this relationship between groups (dyslexia vs. control) or outcome measures (reading vs. spelling). Thus, our results are in line with other studies that do not provide evidence for the relationship between SL and literacy skills (e.g., Nigro et al., 2016; West et al. 2018a; 2018b; Schmalz et al., 2018), despite theoretical claims and experimental evidence of the existence of this relationship from other studies (e.g., Arciuli, 2018; Arciuli & Simpson, 2012; Treiman, 2018; von Koss Torkildsen et al., 2019). Furthermore, our findings highlight the importance of controlling for participant-level variables when investigating the link between SL and literacy attainment.

The absence of evidence for a (strong) relationship between SL and literacy skills in the present study may have a number of explanations. Although these null results may simply be due to chance, several methodological choices may have influenced the outcomes of the present study. More specifically, the SL tasks reported here are not an exact replication of previous studies and we consider a unique range of participant-level variables. Although our VSL task is identical in statistical structure to the task used in previous studies (Arciuli & Simpson, 2012; Qi et al., 2019; von Koss Torkildsen et al., 2019), it involves a different set of stimuli and a novel online measure of learning during exposure (i.e., the task was self-paced, see Siegelman, Bogaerts, Kronenfeld, & Frost, 2018; van Witteloostuijn et al., 2019). Similarly, the SRT task resembles tasks used by Hung et al. (2018) and Van der Kleij et al. (2018), but notable differences include the sequence to be learned (e.g., Hung et al., 2018: a 12-item sequence; here: a 10-item sequence), the number of exposures (e.g., van der Kleij et al., 2018:70 exposures to the sequence prior to the random block; here: 24 exposures prior to the random block), and the visual set-up of the task (e.g., van der Kleij et al., 2018: three locations on the screen presented horizontally; here: four locations presented as a quadrant). Schmalz et al. (2018) and Elleman, Steacy, and Compton (2019) suggested that the variety of *p*-values in the literature examining the association between SL and literacy skills is at least partially due to such methodological choices, the idea being that if the true association is relatively small, it may only appear under certain experimental conditions. These choices could then involve the type of statistical structure tested (e.g., adjacent vs. non-adjacent dependencies), the modality of the task (e.g., visual vs. auditory), the type of task used (e.g., VSL vs. SRT) and the type of instruction given to participants (i.e., more or less implicit). Furthermore, current SL tasks are known to show low correlations among each other, which may help explain differences in detectability when investigating the relationship between SL and other cognitive or linguistic skills (e.g., Schmalz et al., 2018; Siegelman & Frost, 2015).

Another explanation previously put forward is the idea that SL may play a less prominent role in more transparent orthographies than English (such as Dutch, examined here), since grapheme-phoneme correspondences in these orthographies are less complex and thus potentially easier to acquire through explicit instruction (see for example, Elleman et al., 2019; Nigro et al., 2016; Schmalz et al., 2018). This seems a less likely explanation in our opinion, since other studies involving (semi-)transparent orthographies such as Norwegian (Von Koss Torkildsen et al., 2019) and Icelandic (Sigurdardottir et al., 2017) report significant associations between (V)SL tasks and reading performance, even after considering a range of reading-related abilities (Von Koss Torkildsen et al., 2019). Moreover, von Koss Torkildsen et al. report a comparable effect size as found for English (Arciuli & Simpson, 2012), which suggests similar influences of SL on reading performance in (semi-)transparent and opaque orthographies.

Recently, concerns have been raised about the reliability of statistical learning measures (e.g., Kidd, Donnelly, & Christiansen, 2017; Siegelman & Frost, 2015; Siegelman, Bogaerts, & Frost, 2017; West et al., 2018a; 2018b), especially in child participants (Arnon, 2020, 2019b), which limits their appropriateness for studies of individual differences. The statistical learning measures in the present study had split-half reliabilities of $r = 0.71$ (SRT) and $r = 0.70$ and $r = 0.78$ (VSL 2-AFC and 3-AFC, respectively). Previous reports on the reliability of statistical learning measures in children have been less promising, with split-half reliabilities between $r = -0.04$ (ASL) and $r = 0.46$ and $r = 0.59$ on a VSL (Arnon, 2020). In their study of the SRT task, West et al. (2018a) report split-half reliabilities of between $r = 0.17$ and $r = 0.75$. Ideally, the reliability coefficients of psychological measurements reach the value of $r = 0.80$ (e.g., Nunnally & Bernstein, 1994; Streiner, 2003). We could say, therefore, that the reliability coefficients of the statistical learning measures used in the present study approach psychometric standards, although it is important to emphasize that there remains room for improvement.

Finally, we found that there was quite some variation in the VSL performance of children. A substantial proportion of children did not perform above chance-level (even though at the group level, children demonstrated sensitivity to the statistical regularities). The presence of relatively many scores representing around chance-level performance at the individual level within the groups may have hampered finding a relationship between literacy and statistical learning ability.

To clarify the true relationship between SL and literacy acquisition, an important aim for future research is to develop SL tasks that are increasingly reliable and therefore suitable for examining individual differences (e.g., Arnon, 2020; Kidd et al., 2017). Additionally, the present state of the field stresses the need for (exact) replications and large-scale (cross-linguistic) studies, preferably using a fixed set of tasks. We would also like to stress the added value of pre-registration and registered reports, which could help minimize problems such as a publication bias in the field and may thereby clarify the nature of the relationship between statistical learning and literacy skills (e.g., Schmalz et al., 2017; van Witteloostuijn et al., 2017). Theoretical and pedagogical models of reading and spelling should be extended to incorporate SL to enable the formulation of more specific and testable hypotheses for future studies such as 'at what stage of learning to read and spell is SL of importance?' and 'what type of SL is most closely associated with literacy acquisition?' With the accumulation of evidence, meta-analytic analyses may provide insight into the strength of the relationship between SL and literacy skills, which in turn can clarify its relevance for clinical practice and potential use in treatment for individuals with dyslexia. Meta-regression techniques could inform us about potential moderators of the effect such as participant characteristics (e.g., age, native orthography) and methodological choices regarding the SL task (e.g., type of structure, modality).

To conclude, the results of the present study fit with the pattern of varying *p*-values in the field more generally: although we find evidence of correlations between SRT performance and reading and spelling in children with dyslexia (although weak and uncontrolled for potential participant level confounds), no evidence for a relationship between SL and literacy attainment was found once we considered participant-level characteristics such as age, nonverbal reasoning, attention and phonological skills and when we considered the whole sample of children with and without dyslexia. Although these null results may simply be due to chance, it may also suggest that the link between SL and literacy skills may be less strong than previously hypothesized and is likely influenced by methodological choices made in individual studies.

ACKNOWLEDGMENTS

We would first and foremost like to thank all children, their parents and the primary schools (De Bosmark, De Kemp, De Ludgerschool and De Ieme) that participated in our study. We are also grateful to Imme Lammertink for collaboration on the development of the statistical learning measures. Furthermore, we thank Iris Broedelet, Sascha Couvee and Darlene Keydeniers for help with testing the control group and Dirk Jan Vet for technical assistance in developing the tasks. This work was supported by the Netherlands Organization for Scientific Research (NWO) through a personal Vidi grant awarded to Judith Rispens (grant number 276-89-005). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

ENDNOTES

[1] RAN, non-word repetition and digit span forward are grouped together under the label 'phonological skills' for ease of reference. The RAN task requires a complex set of skills: for example, visual recognition, the integration of visual stimuli with stored representations, and the access and retrieval of the associated phonological representations (Norton & Wolf, 2012). Non-word repetition involves existing phonological and lexical representations and phonological short-term memory (Rispens & Baker, 2012), while the digit span assesses phonological short-term memory (e.g., Bull, Espy, & Wiebe, 2008).

[2] Since we were interested in the overall effect of the phonological skill measures, we attempted to summarize these four subtests through maximum likelihood factor analysis. We aimed to reduce all four subtests to one single factor as a minimum of three variables per factor is required (e.g., Tabachnick & Fidell, 2007). However, one factor was deemed insufficient (χ^2 (2) = 13.65, p = .0011). For completeness and transparency, we performed identical confirmatory regression analyses using the single score obtained through factor analysis instead of entering the four phonological skill measures individually (see supplementary analyses on the Open Science Framework). This alternative choice of statistics does not change the main outcomes of the model.

[3] As a reviewer noted, this interaction could reflect a ceiling (or bottom) effect in literacy skills. That is, literacy skills, as a function of group and phonological skills (or SL) could meet a ceiling (or bottom) when they (i.e., literacy skills) get high (or low) enough.

[4] SL may play a more prominent role in pseudo-word reading than in real word reading, due to the fact that pseudo-words have not been encountered before and therefore readers have to read indirectly through grapheme-phoneme mappings (see for example, van der Kleij et al., 2018). Thus, we performed identical confirmatory regression analyses using pseudo-word reading as an outcome measure instead of both reading measures combined (see supplementary analyses on the Open Science Framework). This alternative analysis provides a similar pattern of results. Most importantly, removing the SL measures from the model does not significantly decrease the model's fit (F(8, 162) = 1.244, p = .28).

DATA AVAILABILITY STATEMENT

Summary level data and R Markdown and html files detailing our analyses are available on the Open Science Framework.

ORCID


Merel van Witteloostuijn https://orcid.org/0000-0003-3159-8247

REFERENCES

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5)*. Arlington, TX: American Psychiatric Publishing.

Arciuli, J. (2017). The multi-component nature of statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), 20160058.

Arciuli, J. (2018). Reading as statistical learning. *Language, Speech, and Hearing Services in Schools*, *49*(3S), 634–643.

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. (2011). Statistical learning in typically developing children: The role of age and speed of stimulus presentation. *Developmental Science*, *14*(3), 464–473.

Arciuli, J., & Simpson, I. C. (2012). Statistical learning is related to reading ability in children and adults. *Cognitive Science*, *36* (2), 286–304.

Arnon, I. (2020). Do current statistical learning tasks capture stable individual differences in children? An investigation of task reliability across modality. *Behavior Research Methods*, *52*(1), 68–81.

Arnon, I. (2019b). Statistical learning, implicit learning, and first language acquisition: A critical evaluation of two developmental predictions. *Topics in Cognitive Science*, *11*(3), 504–519.

Braams, T., & de Vos, T. (2015). *Schoolvaardigheidstoets Spelling*. Amsterdam, The Netherlands: Boom uitgevers Amsterdam.

Brus, B., & Voeten, M. (1972). Eén Minuut Test. In *Vorm A en B. Schoolvorderingen voor het lezen, bestemd voor het tweede t/m het vijfde leerjaar van de lagere school*. Nijmegen, The Netherlands: Berkhout Testmateriaal.

Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology*, *33*(3), 205–228.

Capel, D. (2018). *Sequential learning, domain generality, and developmental dyslexia*. Utrecht, The Netherlands: LOT dissertations.

Clark, G. M., & Lum, J. A. (2017). Procedural memory and speed of grammatical processing: Comparison between typically developing children and language impaired children. *Research in Developmental Disabilities*, *71*, 237–247.

De Bree, E., Wijnen, F., & Gerrits, E. (2010). Non-word repetition and literacy in Dutch children at-risk of dyslexia and children with SLI: Results of the follow-up study. *Dyslexia*, *16*(1), 36–44.

De Jong, P. F., & van der Leij, A. (1999). Specific contributions of phonological abilities to early reading acquisition: Results from a Dutch latent variable longitudinal study. *Journal of Educational Psychology*, *91*(3), 450–476.

Elleman, A. M., Steacy, L. M., & Compton, D. L. (2019). The role of statistical learning in word reading and spelling development: More questions than answers. *Scientific Studies of Reading*, *23*, 1–7. https://doi.org/10.1080/10888438.2018.1549045

Fox, J., Weisberg, S., Adler, D., Bates, D., Baud-Bovy, G., Ellison, S., … Heiberger, R. (2012). *Package 'car'*. Vienna, Austria: R Foundation for Statistical Computing.

Frost, R., Siegelman, N., Narkiss, A., & Afek, L. (2013). What predicts successful literacy acquisition in a second language? *Psychological Science*, *24*(7), 1243–1252.

Furnes, B., & Samuelsson, S. (2010). Predicting reading and spelling difficulties in transparent and opaque orthographies: A comparison between Scandinavian and US/Australian children. *Dyslexia*, *16*(2), 119–142.

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.

Hedenius, M., Persson, J., Alm, P. A., Ullman, M. T., Howard, J. H., Jr., Howard, D. V., & Jennische, M. (2013). Impaired implicit sequence learning in children with developmental dyslexia. *Research in Developmental Disabilities*, *34*(11), 3924–3935.

Howard, J. H., Jr., 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.

Hung, Y.-H., Frost, S. J., Molfese, P., Malins, J. G., Landi, N., Einar Mencl, W., … Pugh, K. R. (2018). Common neural basis of motor sequence learning and word recognition and its relation with individual differences in reading skill. *Scientific Studies of Reading*, *23*, 89–100. https://doi.org/10.1080/10888438.2018.1451533

Jiménez-Fernández, G., Vaquero, J. M., Jiménez, L., & Defior, S. (2011). Dyslexic children show deficits in implicit sequence learning, but not in explicit sequence learning or contextual cueing. *Annals of Dyslexia*, *61*(1), 85–110.

Kelly, S. W., Griffiths, S., & Frith, U. (2002). Evidence for implicit sequence learning in dyslexia. *Dyslexia*, *8*(1), 43–52.

Kidd, E., Donnelly, S., & Christiansen, M. H. (2017). Individual differences in language acquisition and processing. *Trends in Cognitive Sciences*, *22*(2), 154–169.

Kort, W., Schittekatte, M., & Compaan, E. (2008). *Clinical Evaluation of Language Fundamentals-4-NL (CELF-4-NL)*. Amsterdam, The Netherlands: Pearson.

Lammertink, I., Boersma, P., Rispens, J., & Wijnen, F. (2020). Visual statistical learning in children with and without DLD and its relation to literacy in children with DLD. *Reading and Writing*, *33*(6), 1557–1589.

Lammertink, I., Boersma, P., Wijnen, F., & Rispens, J. (2020). Children with developmental language disorder have an auditory verbal statistical learning deficit: Evidence from an online measure. *Language Learning*, *70*(1), 137–178.

Le Clercq, C. M., van der Schroeff, M. P., Rispens, J. E., Ruytjens, L., Goedegebure, A., van Ingen, G., & Franken, M. C. (2017). Shortened nonword repetition task (NWR-S): A simple, quick, and less expensive outcome to identify children with combined specific language and reading impairment. *Journal of Speech, Language, and Hearing Research*, *60*(8), 2241–2248.

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.

Menghini, D., Hagberg, G. E., Caltagirone, C., Petrosini, L., & Vicari, S. (2006). Implicit learning deficits in dyslexic adults: An fMRI study. *NeuroImage*, *33*(4), 1218–1226.

Miles, T. R. (2004). Some problems in determining the prevalence of dyslexia. *Electronic Journal of Research in Educational Psychology*, *2*(2), 5–12.

Misyak, J. B., & Christiansen, M. H. (2012). Statistical learning and language: An individual differences study. *Language Learning*, *62*(1), 302–331.

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

Nicolson, R. I., & Fawcett, A. J. (2011). Dyslexia, dysgraphia, procedural learning and the cerebellum. *Cortex*, *47*(1), 117–127.

Nigro, L., Jiménez-Fernández, G., Simpson, I. C., & Defior, S. (2016). Implicit learning of non-linguistic and linguistic regularities in children with dyslexia. *Annals of Dyslexia*, *66*(2), 202–218.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*(1), 1–32.

Norton, E. S., & Wolf, M. (2012). Rapid automatized naming (RAN) and reading fluency: Implications for understanding and treatment of reading disabilities. *Annual Review of Psychology*, *63*, 427–452.

Nunnally, J., & Bernstein, I. (1994). Psychometric theory (3rd ed.). McGraw-Hill.

Plante, E., & Gómez, R. L. (2018). Learning without trying: The clinical relevance of statistical learning. *Language, Speech, and Hearing Services in Schools*, *49*(3S), 710–722.

Pothos, E. M., & Kirk, J. (2004). Investigating learning deficits associated with dyslexia. *Dyslexia*, *10*(1), 61–76.

Psychology Software Tools, Inc. (2012). E-Prime 2.0.

Qi, Z., Sanchez Araujo, Y., Georgan, W. C., Gabrieli, J. D., & Arciuli, J. (2019). Hearing matters more than seeing: A cross-modality study of statistical learning and reading ability. *Scientific Studies of Reading*, *23*, 101–115. https://doi.org/10.1080/10888438.2018.1485680

Raven, J. & Raven, J. (2003). Raven progressive matrices [Measurement instrument]. In R. S. McCallum (Ed.) Handbook of nonverbal assessment (pp. 223–237). New York, United States: Kluwer Academic/Plenum Publishers.

Rispens, J., & Baker, A. (2012). Nonword repetition: The relative contributions of phonological short-term memory and phonological representations in children with language and reading impairment. *Journal of Speech, Language, and Hearing Research*, *55*(3), 683–694.

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.

Schmalz, X., Altoè, G., & Mulatti, C. (2017). Statistical learning and dyslexia: A systematic review. *Annals of Dyslexia*, *67*(2), 147–162.

Schmalz, X., Moll, K., Mulatti, C., & Schulte-Körne, G. (2018). Is statistical learning ability related to reading ability, and if so, why? *Scientific Studies of Reading*, *23*, 64–76. https://doi.org/10.1080/10888438.2018.1482304

Schneider, W., Eschman, A., & Zuccolotto, A. (2012). *E-Prime 2.0 reference guide manual*. Pittsburgh, PA: Psychology Software Tools.

Schittekatte, M., Groenvynck, H., Fontaine, J., & Dekker, P. (2007). TEAch: Test of Everyday Attention for Children, *Dutch version [Measurement instrument]*. Amsterdam, The Netherlands: Pearson.

Siegel, L. S. (2006). Perspectives on dyslexia. *Paediatrics and Child Health*, *11*(9), 581–587.

Siegelman, N., Bogaerts, L., Christiansen, M. H., & Frost, R. (2017). Towards a theory of individual differences in statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), 20160059.

Siegelman, N., Bogaerts, L., & Frost, R. (2017). Measuring individual differences in statistical learning: Current pitfalls and possible solutions. *Behavior Research Methods*, *49*(2), 418–432.

Siegelman, N., Bogaerts, L., Kronenfeld, O., & Frost, R. (2018). Redefining "learning" in statistical learning: What does an online measure reveal about the assimilation of visual regularities? *Cognitive Science*, *42*, 692–727.

Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language*, *81*, 105–120.

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., 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), 1–15.

Snowling, M. J., & Melby-Lervåg, M. (2016). Oral language deficits in familial dyslexia: A meta-analysis and review. *Psychological Bulletin*, *142*(5), 498–545.

Staels, E., & Van den Broeck, W. (2017). A specific implicit sequence learning deficit as an underlying cause of dyslexia? Investigating the role of attention in implicit learning tasks. *Neuropsychology*, *31*(4), 371–382.

Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, *80*(1), 99–103.

Swanson, H. L., & Howell, M. (2001). Working memory, short-term memory, and speech rate as predictors of children's reading performance at different ages. *Journal of Educational Psychology*, *93*(4), 720–734.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*. Boston, MA: Allyn & Bacon/Pearson Education.

Treiman, R. (2018). Statistical learning and spelling. *Language, Speech, and Hearing Services in Schools*, *49*(3S), 644–652.

Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, *134*(4), 552–564.

Vakil, E., Lowe, M., & Goldfus, C. (2015). Performance of children with developmental dyslexia on two skill learning tasks—Serial reaction time and tower of hanoi puzzle: A test of the specific procedural learning difficulties theory. *Journal of Learning Disabilities*, *48*(5), 471–481.

Van den Bos, K. P., & Spelberg, H. (2007). *Continu Benoemen & Woorden Lezen*. Amsterdam, The Netherlands: Boom uitgevers Amsterdam.

Van den Bos, K. P., Spelberg, H., Scheepsma, A., & De Vries, J. (1994). *De Klepel. Vorm A en B. Een test voor de leesvaardigheid van pseudowoorden. Verantwoording, handleiding, diagnostiek en behandeling*. Nijmegen, The Netherlands: Berkhout.

Van der Kleij, S. W., Groen, M. A., Segers, E., & Verhoeven, L. (2018). Sequential implicit learning ability predicts growth in reading skills in typical readers and children with dyslexia. *Scientific Studies of Reading*, *23*, 77–88. https://doi.org/10.1080/10888438.2018.1491582

Van Setten, E. R., Tops, W., Hakvoort, B. E., van der Leij, A., Maurits, N. M., & Maassen, B. A. (2017). L1 and L2 reading skills in Dutch adolescents with a familial risk of dyslexia. *PeerJ*, *5*, e3895. https://doi.org/10.7717/peerj.3895

Van Witteloostuijn, M., Boersma, P., Wijnen, F., & Rispens, J. (2017). Visual artificial grammar learning in dyslexia: A meta-analysis. *Research in Developmental Disabilities*, *70*, 126–137.

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.

Von Koss Torkildsen, J., Arciuli, J., & Wie, O. B. (2019). Individual differences in statistical learning predict children's reading ability in a semi-transparent orthography. *Learning and Individual Differences*, *69*, 60–68.

Wagenmakers, E. J., Wetzels, R., Borsboom, D., van der Maas, H. L., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, *7*(6), 632–638.

West, G., Shanks, D., & Hulme, C. (2018). *Sustained attention, not procedural learning, is a predictor of language, reading and arithmetic skills in children*, *25*, 47–63. https://doi.org/10.31234/osf.io/afrms

West, G., Vadillo, M. A., Shanks, D. R., & Hulme, C. (2018). The procedural learning deficit hypothesis of language learning disorders: We see some problems. *Developmental Science*, *21*(2), e12552. https://doi.org/10.1111/desc.12552

How to cite this article: van Witteloostuijn M, Boersma P, Wijnen F, Rispens J. The contribution of individual differences in statistical learning to reading and spelling performance in children with and without dyslexia. *Dyslexia*. 2021;27:168–186. https://doi.org/10.1002/dys.1678

APPENDIX A

TABLE A1 TD (*N* = 50) Pearson correlations between all measures

	VSL 3-AFC	SRT	Reading words	Reading non-words	Spelling	RAN letters	RAN pictures	NWR	Digit span
VSL 2-AFC	0.71	0.02	0.19	0.07	0.04	−0.07	−0.19	0.18	−0.17
VSL 3-AFC		0.13	0.17	0.08	−0.07	−0.04	−0.08	0.20	−0.19
SRT			−0.17	−0.26	−0.12	0.29	0.17	0.10	−0.06
Reading words				0.81	0.49	−0.62	−0.54	0.32	0.03
Reading non-words					0.39	−0.60	−0.60	0.31	0.02
Spelling						−0.10	−0.09	0.29	0.23
RAN letters							0.52	−0.14	−0.07
RAN pictures								−.14	.07
NWR-S									0.43

Note: Correlations between RAN letters or RAN pictures and other measures are expected to be negative, due to the way RAN is measured (number of seconds spent on the task, that is, higher score is lower performance).

TABLE A2 DD (*N* = 50) Pearson correlations between all measures

	VSL 3-AFC	SRT	Reading words	Reading non-words	Spelling	RAN letters	RAN pictures	NWR	Digit span
VSL 2-AFC	0.55	−0.08	0.04	0.09	0.04	−0.01	−0.28	0.12	0.11
VSL 3-AFC		0.03	0.17	0.10	0.23	0.09	−0.12	0.21	0.07
SRT			0.37	0.26	0.38	−0.23	−0.15	0.07	0.04
Reading words				0.75	0.65	−0.54	−0.38	0.31	0.27
Reading non-words					0.57	−0.51	−0.22	0.37	0.37
Spelling						−0.12	−0.13	0.20	0.22
RAN letters							0.40	−0.26	−0.18
RAN pictures								−0.12	−0.21
NWR-S									0.36

Note: Correlations between RAN letters or RAN pictures and other measures are expected to be negative, due to the way RAN is measured (number of seconds spent on the task, that is, higher score is lower performance).

APPENDIX B

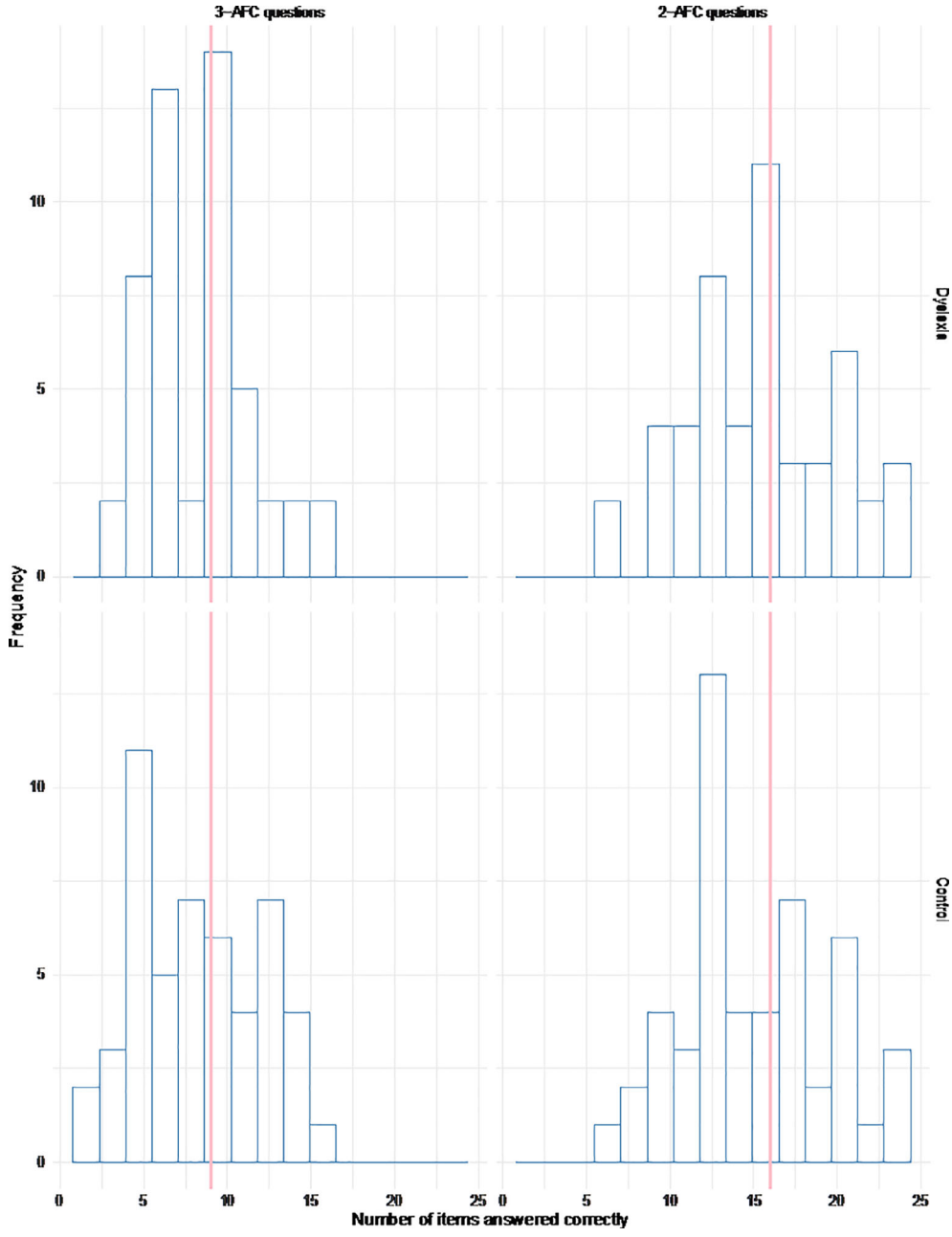

FIGURE B1 Histogram of items correct. The pink lines represent the level at which individuals exceed chance level as determined by the binomial distribution (see Section 3.3.3; 16/24 or higher for 2-AFC and 9/16 or higher for 3-AFC) [Colour figure can be viewed at wileyonlinelibrary.com]

	2-AFC	3-AFC
TD	20/50 = 40% above chance	22/50 = 44% above chance
DD	23/50 = 46% above chance	25/50 = 50% above chance

TABLE B1 Distribution of 'above chance' scores of the individuals within each group for both tasks