Journal of Experimental Child Psychology 246 (2024) 106002

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Statistical learning ability at 17 months relates to early reading skills via oral language

Padraic Monaghan[a,*], Lana S. Jago[a,b], Lydia Speyer[a], Heather Turnbull[c], Katie J. Alcock[a], Caroline F. Rowland[c,d], Kate Cain[a]

[a] Lancaster University, Lancaster LA1 4YF, UK
[b] Liverpool John Moores University, Liverpool L3 5AH, UK
[c] University of Liverpool, Liverpool L69 7ZX, UK
[d] Max Planck Institute for Psycholinguistics, 6525 XD Nijmegen, the Netherlands

ARTICLE INFO

Article history:

Keywords:
Reading development
Early literacy
Oral language
Statistical learning
Artificial language learning
Non-adjacent dependencies

ABSTRACT

Statistical learning ability has been found to relate to children's reading skills. Yet, statistical learning is also known to be vital for developing oral language skills, and oral language and reading skills relate strongly. These connections raise the question of whether statistical learning ability affects reading via oral language or directly. Statistical learning is multifaceted, and so different aspects of statistical learning might influence oral language and reading skills distinctly. In a longitudinal study, we determined how two aspects of statistical learning from an artificial language tested on 70 17-month-old infants—segmenting sequences from speech and generalizing the sequence structure—related to oral language skills measured at 54 months and reading skills measured at approximately 75 months. Statistical learning segmentation did not relate significantly to oral language or reading, whereas statistical learning generalization related to oral language, but only indirectly related to reading. Our results showed that children's early statistical learning ability was associated with learning to read via the children's oral language skills.

* Corresponding author.
E-mail address: p.monaghan@lancaster.ac.uk (P. Monaghan).

https://doi.org/10.1016/j.jecp.2024.106002
0022-0965/

Introduction

Early reading development is predicted by children's oral language skills (Hjetland et al., 2020; Kendeou et al., 2009; Language and Reading Research Consortium [LARRC] & Chiu, 2018; Lee, 2011; Ouellette, 2006; Storch & Whitehurst, 2002). Much of this previous research has focused on the prediction of word recognition and reading comprehension by separate language skills (Gough & Tunmer, 1986), demonstrating that early word reading is influenced by code-related language skills (Catts et al., 2006; Storch & Whitehurst, 2002) and that early reading and listening comprehension are predicted by oral language skills (LARRC & Chiu, 2018; Muter et al., 2004; Silva & Cain, 2015).

The simple view of reading (Gough & Tunmer, 1986; Hoover & Gough, 1990) describes this separation between decoding skills relating to word recognition (Catts et al., 2006; Storch & Whitehurst, 2002), which then relates to reading comprehension, and oral language skills that support reading comprehension (LARRC & Chiu, 2018; Muter et al., 2004; Silva & Cain, 2015). However, more recent accounts of reading development (i.e., both word recognition and reading comprehension) have indicated the interplay between oral language skills and precursors of word recognition (Nation & Snowling, 2004). For instance, oral language skills have been proposed to support not only reading comprehension but also development of letter knowledge and grapheme-to-phoneme correspondences (e.g., Duff et al., 2015; Kim et al., 2014; Ouellette & Beers, 2010; Roth et al., 2002; Share, 2004), which in turn relate to word recognition. Furthermore, different aspects of oral language skills have been shown to affect reading development at different stages. Ouellette and Beers (2010) found that oral vocabulary related to word recognition for older children (Grade 6) and also for younger children (Grade 1), but only for irregular words, but listening comprehension (along with skills relating to decoding written to spoken forms as measured by nonword reading) did predict reading development at both ages, although Ricketts et al. (2016) established this link between oral vocabulary and regular and irregular word recognition for younger children. Both these studies support a view of oral language having a role in word recognition as well as reading comprehension.

This more interactive view of reading was supported by a meta-analysis of 64 studies that investigated the role of oral language skills in reading comprehension development (Hjetland et al., 2020). The authors found that, across this extant literature, letter knowledge and phoneme awareness skills were related to word recognition (note that they did not investigate nonword reading tasks in their meta-analysis), which was in turn related to reading comprehension, and oral language was related to precursors of word recognition as well as to reading comprehension. With age, the relation between oral language and reading comprehension increased in strength, and the relation between oral language and precursors of word recognition decreased in strength. Thus, although code-related skills and oral language skills were separable to a degree, they were found to be inter-related for both word recognition and reading comprehension.

This more nuanced view of precursors of reading skills is consistent with models of reading where word recognition and comprehension skills interpenetrate, such as the triangle model of reading (Harm & Seidenberg, 2004). In this model, representations of sounds, meanings, and written forms of words interact with one another, and connections between representations become stronger with more experience. The role of oral language skills in the reading system can then be implemented by pretraining models on mappings between sound and meaning representations of words, which can be varied by pretraining the model with smaller to larger vocabularies and with lower to higher fidelity of sound and meaning representations (Chang & Monaghan, 2019; Metsala & Walley, 1998). Models with varied oral language skills can then be exposed to written forms of words to determine how oral language skills in the models affect development of reading in terms of both pronunciation and accessing meaning for words. Such computational models are able to forge the link between data-oriented and theoretical models of reading (Chang et al., 2019; Harm & Seidenberg, 2004; Hjetland et al., 2020), testing the causal consequences of precursor skills on reading development.

In summary, the links between oral language skills and early reading development are well-established, with a broad separation of oral language and code-related skills concerning representations and mappings of letters and sounds (such as phoneme awareness and letter naming skills; Hjetland et al., 2020), but with a degree of interactivity among these precursors of reading. However, the ability to *learn associations* among written, spoken, and meaning representations of words—in other words, children's statistical learning ability—has been proposed as an individual difference that can explain variation in both oral language skills (Christiansen, 2019; Frost et al., 2015; Kidd et al., 2018) and reading development (Arciuli & Simpson, 2012; Frost, 2012; Frost et al., 2013).

There are two possible ways in which statistical learning may highlight the relation between oral language and reading development. It could be that statistical learning is implicated in development of oral language, and then oral language skills relate to children's early reading skills. In this case, statistical learning would relate to reading only via oral language skills. Alternatively, statistical learning may have a direct influence on reading in addition to a potentially mediated effect on reading via oral language.

However, statistical learning covers a panoply of processes over different kinds of structures (Bogaerts et al., 2021; Frost et al., 2015; Kidd et al., 2018), from associations between co-occurring stimuli (Cunillera et al., 2010; Perruchet & Vinter, 1998), to learning transitions between sequences (Oliveira et al., 2023; Saffran, 2003; West et al., 2018), to clustering stimuli that occur in similar contexts (Mintz, 2002; St. Clair et al., 2010). In particular, statistical learning has been related to possibly distinct processes of oral vocabulary and grammar acquisition (but see Frost & Monaghan, 2016, and Perruchet et al., 2004, for discussion). These distinct processes have distinguished between chunking or identifying regular sequences in language (e.g., Marchetto & Bonatti, 2015; Peña et al., 2002; Saffran, 2003) and identifying more complex structures over which generalization can be performed (Wang et al., 2019; Wilson et al., 2020). Thus, it is necessary to clarify which aspects of statistical learning are in question with respect to development of both oral language and literacy skills.

Children's oral language skills are underwritten by the ability to extract statistical patterns present within language (Bogaerts et al., 2021; Oliveira et al., 2023; Ren et al., 2023). Such statistical learning skills are important for determining the regularities of speech both for determining phonotactic constraints within a language and for extracting words from utterances (Saffran, 2003). For instance, Saffran et al. (1996) demonstrated that young children were sensitive to sequences of syllables that frequently co-occurred together, responding to the changes in transitional probabilities between syllables. Similarly, words that frequently occur, and that can occur with a range of other words such as "the" and "and," provide points of low transitional probability to the next word (Cunillera et al., 2010; Frost et al., 2019), and young children have been shown to be sensitive to this kind of statistical information (Bortfeld et al., 2005). Once a word is identified, statistical associations between the word and potential referents in the environment are required to determine meaning (Roembke et al., 2023). For learning grammar, young children are sensitive to words that occur together when they are separated by intervening words that vary (Gómez & Gerken, 1999; Marchetto & Bonatti, 2015), reflecting statistical processing involved in grammatical processing.

Despite this substantial work unpacking the statistical processes involved in language acquisition, studies that have related statistical learning tasks to oral language skills in children are limited (Gerbrand et al., 2022). Frost et al. (2019), in an individual differences analysis, showed that ability to use statistical learning in order to segment an artificial language at 17 months predicted children's vocabulary development up to 30 months. Relatedly, Evans et al. (2009) found, in a synchronous study, that segmentation of an artificial language in children aged 6 years 5 months to 14 years 4 months correlated with receptive vocabulary. In terms of applying statistical learning to generalize an artificial language, Monaghan et al. (2023) found that this ability at 17 months predicted children's oral language skills at 54 months, and this was independent of children's working memory ability. In a meta-analysis, Ren et al. (2023) found that there was a moderate correlation between statistical learning and oral language skills (see also Isbilen & Christiansen, 2022), which was stronger for infants and adults but slightly weaker for older children (see also Forest et al., 2023). Furthermore, the link between statistical learning tasks that were based on linguistic materials and oral language skills was marginally stronger than that between statistical learning tasks that were non-linguistic (such as visual stimulus sequences) and oral language skills. Ren et al. (2023) also distinguished statistical learning tasks that related to segmentation of stimuli from those that expressed grammatical relations among elements of sequences. They found that oral language skills were significantly related to statistical learning as measured by segmentation tasks, but there was a high degree of variance for grammatical relations tasks relating to natural language skills, such that the relationship was not significant.

There are numerous points where statistical learning ability might also be involved in reading development (see, e.g., Chetail, 2017; Frost, 2012; Siegelman, Rueckl, Lo, et al., 2022; Siegelman, Rueckl, van den Bunt, et al., 2022; Smith et al., 2021), for instance, in acquiring relations among letters and learning to associate letters or sets of letters with sounds and meanings (Lelonkiewicz et al., 2020, 2023). Thus, there are possibilities for similar statistical learning tasks as for oral language skills to also be related to reading, although there is also the possibility that visual statistical learning tasks may more closely relate to reading skills. There is variation in whether a link between statistical learning tasks and reading skills has been found (Lammertink et al., 2020). Arciuli and Simpson (2012) tested 9-year-old children's ability to learn sequences of frequently occurring shapes and found that this related to ability to read single words, a finding that was replicated with sound sequences (Qi et al., 2019) and across languages (von Koss Torkildsen et al., 2019). However, Schmalz et al. (2018) did not find this relationship, and neither did West et al. (2018), who tested spatial sequence statistical learning and reading ability in 7- and 8-year-old children.

As summarized in van Witteloostuijn et al. (2021), the substantial variation in the studies that have attempted to link statistical learning to reading may be due to variation in the type of tests used or the extent to which other mediating or moderating factors are confounded with the tests. Ren et al.'s (2023) meta-analysis considered relations between statistical learning and reading skills and found a similar moderate correlation as for the relation between statistical learning and oral language skills. The size of the effect was moderated by age, such that adults showed a stronger relation than children, which might indicate that greater exposure to statistical relations among letters and sounds may draw more heavily on statistical learning skills with experience (see also Siegelman, Rueckl, Lo, et al., 2022; Siegelman, Rueckl, van den Bunt, et al., 2022). However, there was no difference in the size of the relationship when visual and auditory statistical learning tasks were distinguished, suggesting that there is overlap in the statistical processes involved in oral language and reading skills.

Yet, despite the potential overlap, one of the key potential mediators that is seldom controlled in studies comparing statistical learning with reading is oral language skills. Indeed, when oral language was also considered alongside literacy, West et al. (2018) found that spatial sequence statistical learning *did* relate to oral language, as measured by a test of grammatical processing (the Test for Reception of Grammar–Version 2 [TROG-2]; Bishop, 2003), although statistical learning did not relate directly to reading.

In a preregistered analysis (https://osf.io/p8nvc/?view_only=926a0087b80e419086e48e8959c1a210), we tested the extent to which statistical learning abilities in infancy predicted children's early stages of reading development and whether the prediction was direct or mediated via children's oral language skills. Our study advances the field in three key ways. First, we included measures of different types of statistical learning—learning segmentation (i.e., acquiring knowledge of frequently co-occurring syllables) and learning generalization (i.e., acquiring the structure of the language and abstracting this to accommodate new items consistent with the language structure). Previous studies have tended to focus on only one type of statistical learning test. Second, we measured statistical learning ability as a precursor in time to the measures of language development. Previous studies have largely focused on synchronous investigations of statistical learning and oral language and reading skills, where the nature of the diachronic influence of statistical learning on these language skills cannot be clearly determined. Measuring statistical learning abilities early in children's development can better indicate how they contribute to the construction of children's language skills. Third, we investigated the link between statistical learning and both oral language and reading, to unpack what aspects of oral language are contributing to reading development, by uncovering whether statistical learning processes are common to both or whether there is specificity in these relationships.

Method

Participants

A total of 90 families were initially recruited to participate in a longitudinal study in North West England, investigating language development over the first 6 years of children's lives. Children were tested on a number of measures between 9 and 54 months (Language 0–5 Project; (Peter et al., 2019; Rowland et al., 2024). At 54 months, families were invited to continue with the longitudinal study for a further 2 years, and 33 families volunteered to participate. All children were monolingual, were born at term, had normal vision and hearing, and had receptive vocabulary within the normal range when tested at 17 months. There were 71 children tested at 17 months (40 girls and 31 boys aged 16.5–17.5 months; mean age = 517 days). There were 62 to 71 children tested for the measures at 54 months (36 girls and 35 boys aged 53.9–55.3 months), and 31 or 32 children were tested for the measures in Grade 2 school (70+ months: 18 girls and 14 boys aged 74.0–86.5 months).

The tests at 17 and 54 months are described in detail in Monaghan et al. (2023), and so we provide only an outline of those tests here.

Statistical language learning task at 17 months

Materials

To measure statistical language learning, we measured children's ability to detect and use non-adjacent dependencies from continuous artificial speech. Words in the language comprised four sequences where the first and third syllables always co-occurred and the medial syllable varied over two possible syllables (i.e., *bamuso, bagaso, limufe,* and *ligafe*). All speech stimuli were synthesized using the Festival speech synthesizer (Black et al., 1990).

To test statistical learning segmentation, the target words were the words in the language and the foils were three-syllable sequences that occurred during exposure but did not respect the non-adjacent dependency (*sobamu, feliga, gasoli,* and *mufeba*).

To test statistical learning generalization, new syllables were interposed between the non-adjacent dependent syllables as target stimuli (*baniso, baposo, lidufe,* and *livefe)*, and for foil stimuli the novel syllable was used at the beginning or ending of two-syllable sequences that occurred during exposure (*posoba, nifeli, solive,* and *febadu*).

Procedure

Tests were given in the same order for all children to minimize individual variation that is noise and not associated with children's task performance (Cooper et al., 2017; Panter et al., 1992).

For exposure, children listened to the 15-min training speech comprising continuous words concatenated from the language while playing nonverbally with the experimenter. For testing, children were positioned in front of a computer screen with eye movements recorded using an EyeLink 1000 + eye-tracker (SR Research, Ottawa, Ontario, Canada) with speakers positioned to the left and right sides of the screen.

Infants were habituated to a test stimulus occurring to the left or right of the screen, played repeatedly. Trials were halted when children looked away from a picture on the screen for more than 2 s or when 65 s had elapsed, whichever was earlier. Targets and foils were each presented twice. The segmentation test trials were presented first and then the generalization test trials, separated by children watching a 2-min cartoon.

Analysis and scoring

The measure of statistical language learning for each test was the effect size of children's looking preference to the target over the foil stimuli, developed by Frost et al. (2020). Positive values indicate preference for looking to targets over foil stimuli, negative values indicate preference for looking to foils over target stimuli, and values close to zero indicated no preference.

Oral language tasks at 54 months

Materials

We measured oral language using a range of measures testing both oral vocabulary and grammar skills, ranging over receptive and expressive measures. The British Picture Vocabulary Scale–Third Edition (BPVS-3; Dunn et al., 2009) assesses receptive vocabulary, for which children are asked to select one picture (from an array of four pictures) to match a word spoken aloud by the assessor. The Renfrew Word Finding Vocabulary Test (Renfrew & Mitchell, 2010) is a measure of expressive vocabulary and requires children to name a picture. Both can be regarded as measures of vocabulary breadth in that they are proxy measures for the number of words known. The TROG-2 (Bishop, 2003) is a measure of receptive grammar for which children are asked to select one picture (from an array of four pictures) to match a sentence spoken aloud by the assessor. The Renfrew Bus Story–Revised Edition (Renfrew, 2010) includes a story retelling component, and productions can be scored for grammatical complexity (sentence length and use of subordinate clause) to provide a measure of expressive grammar.

Procedure

Children were tested on a wide range of cognitive and language tests in a series of testing sessions conducted in a quiet room with their caregivers present. The four oral language measures included in the current study were tested in this order: (1) the TROG-2, (2) the Renfrew Word Finding Vocabulary Test, (3) the Renfrew Bus Story task, and (4) the BPVS-3.

Analysis and scoring

For the BPVS-3, the Renfrew word finding vocabulary test, and the TROG-2, we used the raw scores. For the Renfrew Bus Story task, we scored both sentence length score and subordinate clause score to reflect language complexity of children's productions.

Reading tasks at 70+ months

Materials

To test word reading skills, children completed two subtests of the Test of Word Reading Efficiency (TOWRE-2; Torgesen et al., 2012): the Sight Word Efficiency subtest (TOWRE-SWE), which tested how many real words could be read accurately in a specified time, and the Phonemic Decoding Efficiency subtest (TOWRE-PDE), which measured accuracy of nonword reading or in a specified time. Thus, these subtests are considered measures of reading efficiency. In addition, we assessed the accuracy and fluency of reading words in context, as well as reading comprehension, using form A of the Passage Reading subtest of the York Assessment of Reading for Comprehension (YARC; Snowling et al., 2009).

Procedure

Children were tested once they all had begun Year 2 of school (ages 6–7 years, second year post-kindergarten in the U.K. school system, Grade 1 in United States), so that their formal schooling was at a similar level (if we had tested children at the same age, given that their birthdays spanned two school year cutoffs, children's formal learning would have differed). Due to COVID-19 restrictions that were in place at the time, children were tested individually in an online call with the experimenter, with paper-based materials sent out to the families by post ahead of the call.

For the YARC, all children read the beginner and Level 1 passages. If they made fewer than 16 errors on the Level 1 passage, they progressed to the Level 2 passage. Then, for the TOWRE-SWE and TOWRE-PDE, to mitigate against potential lagging that arises over video calls, we increased the time children were given to read each word list from 45 s (as per the manual instructions) to 55 s. This was done to account for any delay between when the researcher instructed children to begin reading and when they responded.

Analysis and scoring

For the TOWRE-SWE and TOWRE-PDE, we used age-standardized scores. For the YARC, scores were calculated across two passages: the beginner and Level 1 passages for children who did not read the Level 2 passage and the Level 1 and Level 2 passages for children who did read the Level 2 passage. We followed this design because there was a wide range of reading ability and ages across this sample of children. Scoring protocols in the manual were followed to provide three scores: reading accuracy and reading rate (YARC-accuracy and YARC-rate), which were taken as measures of reading fluency, and reading comprehension (YARC-comprehension). Only Level 1 and Level 2 passages are timed, and therefore reading rate was calculated for the Level 1 passage only for any children who did not read the Level 2 passage.

Results

Preregistration of the analysis and data are available on the Open Science Framework (https://osf.io/p8nvc/?view_only=926a0087b80e419086e48e8959c1a210).

Descriptive statistics

Descriptive statistics for all the measures are shown in Table 1, and the correlations among the measures are shown in Table 2. Table 1 also reports the standardized scores for the BPVS-3 and TROG-2 tests.

Confirmatory factor analysis, structural equation modeling, and mediation analyses were conducted using "lavaan" in R (Rosseel, 2012). We used full information maximum likelihood to account for missing values in the dependent variables, and we also tested models where missing values for predictor variables were also estimated. For the models, we used standard fit indices with cutoff points recommended by Hu and Bentler (1999)—namely root mean square error approximation (RMSEA) < .06, comparative fit index (CFI) > .95, and Tucker–Lewis index (TLI) > .95—and we also employed χ^2 difference tests to determine whether the model approximates the data effectively (where a significant value indicates divergence from the data).

According to our preregistered analysis plan, we first replicated the structural equation model of the relations between the 17-month statistical learning measures and the 54-month oral language measures as reported in Monaghan et al. (2023). We then determined the relations among the reading measures at 70+ months, using confirmatory factor analysis, before combining the 17-month and

Table 1
Descriptive statistics for all measures.

Measure	*N*	Mean	*SD*	Min	Max
Statistical learning at 17 months)					
Segmentation	70	−0.30	0.49	−1.41	0.76
Generalization	61	0.36	0.74	−2.18	1.88
Oral language measures at 54 months					
BPVS-3	71	78.89	11.61	47	106
Renfrew vocabulary	72	33.25	4.32	23	42
TROG-2	72	7.10	3.70	1	17
Renfrew subordinate clause	62	0.98	1.11	0	4
Renfrew sentence length	62	8.43	2.40	4.40	13.80
Reading measures at 75 months					
TOWRE-SWE	32	121.3	15.2	93.0	145.0
TOWRE-PDE	32	118.6	13.5	95.0	145.0
YARC-accuracy	31	114.8	9.0	88.0	125.0
YARC-rate	31	111.4	13.4	78.0	129.0
YARC-comprehension	31	115.6	7.5	101.0	128.0

Note. BPVS-3, British Picture Vocabulary Scale–Third Edition; TROG-2, Test for Reception of Grammar–Version 2; TOWRE, Test of Word Reading Efficiency; SWE, Sight Word Efficiency subtest; PDE, Phonemic Decoding Efficiency subtest; YARC, York Assessment of Reading for Comprehension.

Table 2
Correlations between the statistical learning and oral language measures with the reading measures.

	TOWRE-SWE	TOWRE-PDE	YARC-accuracy	YARC-rate	YARC-comp
Statistical learning segmentation	.02	.17	.06	−.02	.10
Statistical learning generalization	−.13	.06	.19	−.11	.13
BPVS-3	.47**	.40*	.48**	.51**	.33
Renfrew vocabulary	.42*	.38*	.49**	.56**	.53**
TROG-2	.30	.23	.27	.39*	.41*
Renfrew subordinate clause	.42*	.47**	.22	.46*	.18
Renfrew sentence length	.40*	.35	.24	.40*	.23
TOWRE-PDE	.81***				
YARC-accuracy	.81***	.74***			
YARC-rate	.92***	.72***	.77***		
YARC-comp	.71***	.58***	.70***	.73***	

Note. TOWRE, Test of Word Reading Efficiency; SWE, Sight Word Efficiency subtest; PDE, Phonemic Decoding Efficiency subtest; YARC, York Assessment of Reading for Comprehension; BPVS-3, British Picture Vocabulary Scale–Third Edition; TROG-2, Test for Reception of Grammar–Version 2.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

54-month data with the 70+-month data in two ways: first using structural equation modeling to determine pathways from statistical learning tests and the oral language latent variable to reading and then using mediation analysis to determine whether the statistical learning tests related to reading ability directly or indirectly via oral language.

Step 1: Structural equation model replication from Monaghan et al. (2023).

We first replicated the structural equation model of the paths from the statistical learning segmentation and generalization tests at 17 months and a single latent variable of oral language skills at 54 months to which the measures of vocabulary and grammar contributed. The model was a good fit, RMSEA = .043, CFI = .987, TLI = .989, $\chi^2(12) = 13.326$, $p = .346$. with a significant path from statistical learning generalization to oral language, standardized coefficient = −.330, $z = -2.197$, $p = .028$. Fig. 1 shows the resulting model, indicating the paths between the statistical learning measures and the latent variable of oral language.

Step 2: Confirmatory factor analysis: Determining relations among the reading measures

In accordance with our preregistered analysis plan, we next determined the factor structure of the reading measures. We tested whether a model with a factor for the measures of word and nonword reading accuracy and fluency (TOWRE-SWE, TOWRE-PDE, YARC-accuracy, and YARC-rate) with reading comprehension (YARC-comprehension) as a covariate to the latent variable of the other measures of reading (i.e., accuracy and fluency) provided a good fit to the data. There was a good fit to the data, RMSEA = .057, CFI = .996, TLI = .993, $\chi^2(5) = 5.513$, $p = .356$. Fig. 2 shows the results of this confirmatory factor analysis, highlighting paths between the latent variable of reading (comprising both reading fluency and reading comprehension measures) and the individual measures of reading skill. Note that because the reading comprehension measure was a covariate with the latent variable of reading fluency, a model with a single latent variable—reading, combining both the reading fluency measures and the reading comprehension measure—provides the same fit to the data. Thus, the reading measures were coherently described either in terms of one factor—reading—or two factors: reading fluency and reading comprehension. We tested both of these solutions for the reading measures in the next step.

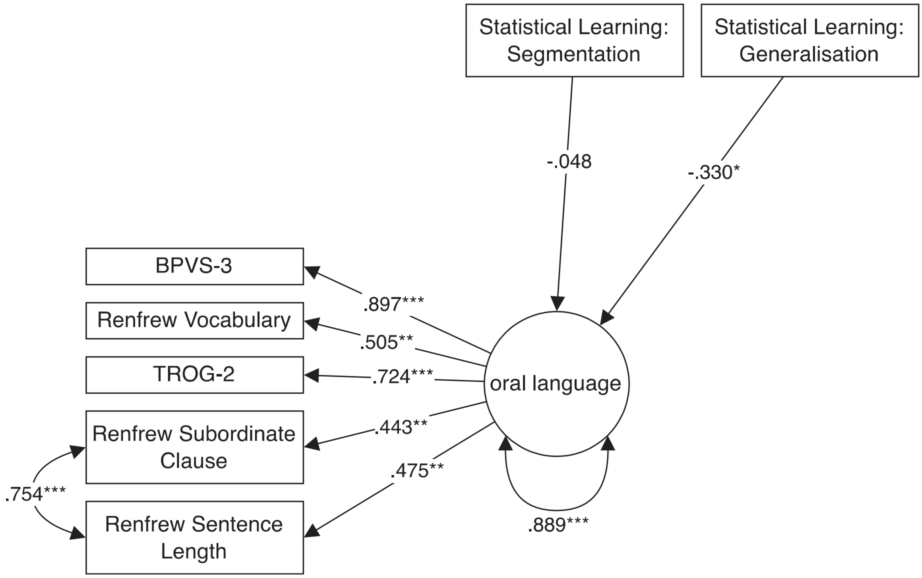

Fig. 1. Relations between 17-month tests of statistical learning segmentation and generalization and general natural language skills at 54 months. Values on paths are standardized estimates. *p < .05; **p < .01; ***p < .001. BPVS-3, British Picture Vocabulary Scale–Third Edition; TROG-2, Test for Reception of Grammar–Version 2.

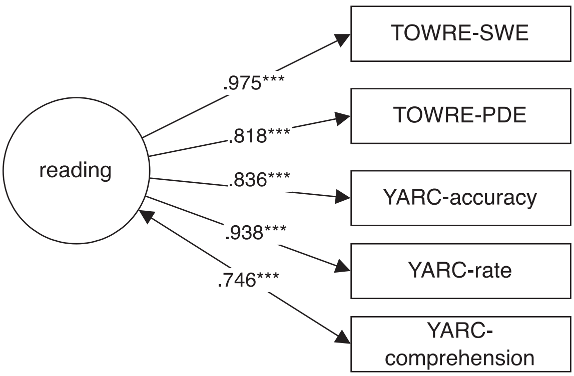

Fig. 2. Confirmatory factor analysis of the reading measures at 70+ months. Values on paths are standardized estimates. ***p < .001. TOWRE, Test of Word Reading Efficiency; SWE, Sight Word Efficiency subtest; PDE, Phonemic Decoding Efficiency subtest; YARC, York Assessment of Reading for Comprehension.

Step 3: Structural equation model: Linking statistical learning at 17 months and oral language skills at 54 months to reading at 70+ months

We first tested a structural equation model with paths from segmentation and generalization artificial learning and oral language onto the reading fluency latent variable and reading comprehension variable, using case-wise maximum likelihood for estimating exogenous covariates (i.e., the *fiml.x* option in lavaan). The structure of the model for the 17-month and 54-month measures and paths between them were the same as those in Step 1, the structure of the model for the reading latent variable and the reading comprehension variable at 70+ months was as examined in Step 2, and there were paths included from the statistical learning and oral language measures to the reading fluency and reading comprehension measures in order to test whether there were direct effects of statistical learning and oral language onto reading skills. The model was not a good fit to the data, according to the TLI criterion, RMSEA = .056, CFI = .960, TLI = .943, $\chi^2(46) = 57.687$, $p = .116$.

We addressed this lack of good fit, in accordance with our preregistered analysis plan, by determining whether a structural equation model as described above but with a single latent variable measure of reading (combining both reading fluency measures and reading comprehension) resulted in a better fit. The model fit was good, RMSEA = .054, CFI = .961, TLI = .949, $\chi^2(49) = 60.272$, $p = .130$, except that TLI was just below the.95 threshold. We note that Bentler (1990) reported that the TLI can fall under the threshold for smaller sample sizes when the other fit measures are met, and so we report this model as our final model. The model is shown in Fig. 3, indicating the relations between oral language and reading, the relations between statistical learning and oral language, and the lack of significant relations between the statistical learning measures and reading (where the latent variable reading represents a combination of reading fluency and reading comprehension tasks).

We repeated this model using case-wise maximum likelihood only for observed variables (and not for exogenous covariates, so using the *fiml* option in lavaan) and found very similar structural relations, although the model failed to meet the criteria for good fit: RMSEA = .112, CFI = .849, TLI = .799, $\chi^2(49) = 86.507$, $p = .001$. See Supplementary Information S1 for the model in the online supplementary material.

The model results indicate that oral language is a strong predictor of reading, as predicted. The results also show that statistical learning generalization is a predictor of oral language, as it was in the model in Step 1. There was no evidence, however, for statistical learning segmentation or generalization directly predicting reading skills at 70+ months. To further test this, we repeated the structural equation model above but removed first the statistical learning segmentation path, then the statistical learning generalization path, and then both and compared these models with the original that contained these paths using log likelihood tests. There was no evidence that removing any of these paths resulted in a poorer fit to the data: removing segmentation path (χ^2 change = 0.077, $p = .782$); removing generalization path (χ^2 change = 0.080, $p = .777$); removing both paths (χ^2 change = 0.125, $p = .939$).

It is possible that relations between statistical learning and reading fluency may be clearer than those between statistical learning and reading comprehension because decoding skills involving learning the associations between letters and sounds may be more dependent on broader statistical

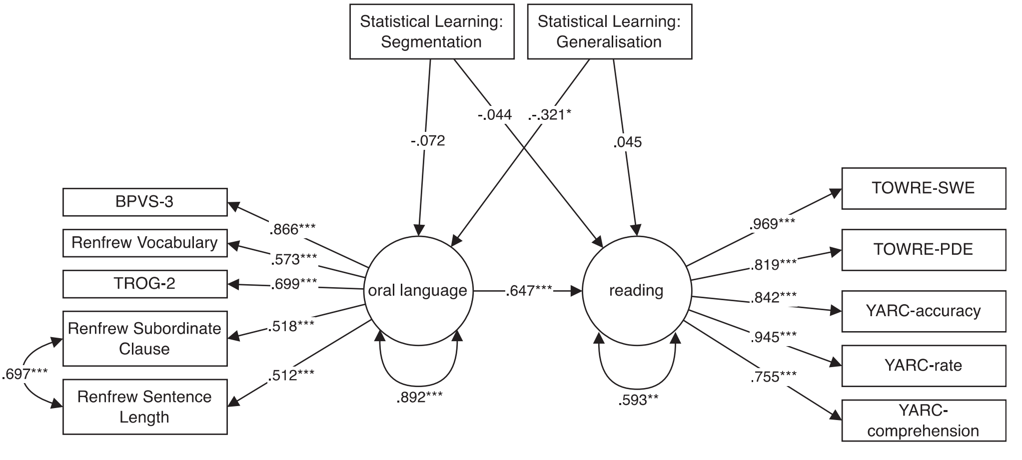

Fig. 3. Structural equation model linking statistical learning measures to oral language and reading. Values on paths are standardized estimates. *$p < .05$; **$p < .01$; ***$p < .001$. BPVS-3, British Picture Vocabulary Scale–Third Edition; TROG-2, Test for Reception of Grammar–Version 2; TOWRE, Test of Word Reading Efficiency; SWE, Sight Word Efficiency subtest; PDE, Phonemic Decoding Efficiency subtest; YARC, York Assessment of Reading for Comprehension.

learning processes (Chetail, 2017; Kemény & Lukács, 2021; Lelonkiewicz et al., 2023). Hence, also including reading comprehension may have obscured potential links between statistical learning and reading given that reading comprehension is known to relate more closely to oral language skills (Hjetland et al., 2020). To test this, we repeated the structural equation model but omitted the measure of reading comprehension from the analysis. Thus, the reading latent variable represented reading fluency without reading comprehension. The significant paths in the model were similar to those from the original model; statistical learning generalization related significantly to oral language, oral language related significantly to reading fluency, but there was no evidence for relations between statistical learning and reading fluency. Note that this analysis was not included in our preregistration, and so we present the resulting model in Supplementary Information S2.

Step 4: Mediation analysis of statistical learning predicting reading via oral language

To ascertain whether there was evidence of direct or indirect pathways from statistical learning measures via oral language to reading, we conducted mediation analyses based on the structural equation model reported in Step 3. We defined direct and indirect pathways from statistical learning measures to reading and estimated their values from the structural equation model. The direct effect of statistical learning segmentation on reading was not significant, (standardized) estimate = −.044, p = .782, and the direct effect of statistical learning generalization on reading also was not significant, estimate = .045, p = .777. The indirect effect of statistical learning segmentation via oral language on reading also was not significant, estimate = −.046, p = .573, but the indirect effect of statistical learning generalization approached significance, estimate = −.208, p = .052. Thus, as with the analyses in Step 3, there was no evidence of a direct influence of statistical learning on reading skills, no evidence for an effect of indirect influence of segmentation on reading, and some evidence that statistical learning generalization may have influenced reading but only indirectly via oral language.

Discussion

Oral language relates strongly to reading development (Duff et al., 2015; Hjetland et al., 2020; Kendeou et al., 2009; LARRC & Chiu, 2018; Muter et al., 2004; Ouellette, 2006; Silva & Cain, 2015). Our study questions the extent to which this relationship can be explained by the role of statistical learning as an ability that underpins both oral language skills (Evans et al., 2009; Frost et al., 2019; Monaghan et al., 2023) and reading (Arciuli, 2018; Chetail, 2017; Frost, 2012; Siegelman, Rueckl, Lo, et al., 2022; Smith et al., 2021). We investigated the extent to which statistical learning ability, as a precursor to oral language, was also critical for learning to read and, if so, whether this relationship was mediated via oral language or was an additional contributor to reading.

The results demonstrated a link between statistical learning ability tested when children were 17 months and oral language skills tested at 54 months. We also found a strong relationship between oral language skills at 54 months and reading skills at 70+ months. We found no evidence of a direct influence of statistical learning on reading skills, but there was some evidence that statistical learning did affect reading skills indirectly via oral language. Thus, statistical learning ability is confirmed as a fundamental contributor to children's oral language development (Isbilen & Christiansen, 2022; Ren et al., 2023) and seems to exert an influence on reading indirectly through these foundational skills.

Our study provides a longitudinal analysis of the unfolding effects of statistical learning ability on later oral language skills at preschool and the effect of these oral language skills on children's learning to read in school. This pattern of results is consistent with a previous synchronous study of statistical learning relating to reading that has also investigated oral language. West et al. (2018) showed that a measure of spatial sequence statistical learning related to oral language but not to reading, although in their study the indirect effect of statistical learning on reading via oral language was not tested and the study was not conducted longitudinally.

Our finding of an indirect influence on reading via oral language is also consistent with studies of relationships between statistical learning and reading in that it provides a possible explanation for why the relationship is sometimes found and sometimes is not found (e.g., Arciuli & Simpson, 2012;

Schmalz et al., 2018). If oral language skills vary between participants, this could be a substantial confound affecting observations of the (indirect) link between statistical learning ability and reading (see also Lammertink et al., 2020; van Witteloostuijn et al., 2021). The current findings, together with those of West et al. (2018), support this conclusion and elucidate the mechanism by which statistical learning influences early reading development.

Our study also aimed to provide answers in three more specific areas of investigation. First, we investigated how distinct measures of statistical learning might explain different aspects of the link between oral language and reading (Bogaerts et al., 2021). Our first measure, statistical learning segmentation, tested children's ability to identify frequently occurring sequences in speech that conform to the statistical structure of the language. In this respect, the test is similar to that used in studies that have investigated statistical learning of co-occurring sequences (e.g., Arciuli & Simpson, 2018; West et al., 2018) and found to relate to oral language skills in Ren et al.'s (2023) meta-analysis. Our second measure, statistical learning generalization, measured children's ability to abstract over the structure of the language to new sequences consistent with the structure. It was this latter measure, however, that related most closely to oral language and indirectly to reading.

In terms of oral language, as language skills develop, the type of statistical relations that support learning may become more sophisticated; learning more advanced vocabulary and grammatical relations, for instance, not only requires acquiring co-occurrences between words and potential referents in the environment that may be sufficient for early vocabulary but also requires abstraction over structures (Gleitman et al., 2005; Kidd, 2012; Monaghan et al., 2023; Oliveira et al., 2023). For reading, as noted by Chetail (2017), learning to read requires acquisition of relations among sometimes spatially distinct letters in speech (e.g., learning that a consonant and an "e" at the end of a word affects the pronunciation of the preceding vowel), and this kind of generalization of the structure is similar to the non-adjacent dependency learning required in our statistical learning generalization test. Treating statistical learning not as a single conglomerate measure but as a suite of potentially distinct mechanisms and abilities (Frost et al., 2015; Kidd et al., 2018) enables a more nuanced view in which aspects of oral language skills contribute to reading development.

The second area of specific investigation was to determine whether examining the relations among statistical learning, oral language, and reading longitudinally reflected the patterns that have been inferred from studies with synchronous measurement. The results showed that statistical learning ability very early in children's development was able to predict oral language skills, and reading as well, but only indirectly via oral language. Hence, correlational studies that show links between oral language skills and statistical learning (Lammertink et al., 2020; Ren & Wang, 2023; Ren et al., 2023) are consistent with the longitudinal effects of statistical learning on oral language that we have shown here. Less certain is the consistency of our results with correlational studies that show links between reading and statistical learning. This relates to our third specific area of investigation, which was to include measures of both oral language and reading in investigating how statistical learning affects each. Our results suggest that at least some aspects of statistical learning processes are common to both oral language and reading but that links between statistical learning and reading should be tested in mediation analyses involving oral language measures, and they should employ measures of different kinds of statistical learning structures. Then, theories of the way in which learning mechanisms resulting in oral language skills relate to reading development can be refined.

An alternative explanation for the absence of evidence for the link between statistical learning and reading is that—from the manifold possibilities of measures of statistical learning—other statistical learning abilities not measured in our study may also be influencing oral language skills and reading and the interface between them. Our statistical learning measures involved acquisition of non-adjacent dependencies, and although sensitivity to this kind of dependency has been shown to distinguish children with different oral language skills (Lammertink et al., 2020), studies focusing on reading skills have rather investigated adjacent dependencies (e.g., Arciuli & Simpson, 2012; Lelonkiewicz et al., 2020, 2023; West et al., 2018; van Witteloostuijn et al., 2021). Non-adjacent dependencies tend to be acquired later in children's development (Marchetto & Bonatti, 2015), and these may relate to different aspects of learning to read than the ability to detect and use adjacent dependencies (Chetail, 2017). Another possibility is that reading is underwritten by different modalities of statistical learning that are more focused on visual statistical learning (Arciuli, 2018; Frost et al., 2015; Raviv &

Arnon, 2018; van Witteloostuijn et al., 2021). Learning statistical relations among visual symbols (i.e., letters) may more directly implicate modality-specific statistical learning abilities in the visual modality rather than the auditory modality (Frost et al., 2015), highlighting again how distinct statistical learning abilities might mediate the relationship between oral language and reading development. Thus, given that our statistical learning tasks used auditory stimuli, this may have reduced our detection of links between visual statistical learning and reading skills (e.g., Lelonkiewicz et al., 2023). However, Ren et al.'s (2023) meta-analysis did not find a different effect of statistical learning task modality relating to reading skills, such that both auditory and visual statistical learning tasks predicted a similar moderate correlation with reading.

Age may also have been a limiting factor in terms of determining the independent contribution of statistical learning skills to reading development. Links between statistical learning and both oral language and reading skills are stronger for infants and adults compared with slightly older children (Ren et al., 2023), and so determining the link between statistical learning and reading skills may grow stronger as children's reading skills develop further. The extent to which this may express statistical learning ability independently of oral language skills remains an open question, however, because with age the link between oral language skills and reading comprehension becomes stronger (Hjetland et al., 2020; Ouellette & Beers, 2010), potentially limiting the possibility for additional variance to be explained by statistical learning measures that are not mediated via oral language.

In terms of the different aspects of oral language relating to reading development, our results are consistent with accounts that posit an interactivity between decoding skills and oral language skills relating to both word recognition and reading comprehension (Chang & Monaghan, 2019; Hjetland et al., 2020; Kim et al., 2014; Ouellette & Fraser, 2009; Ricketts et al., 2016; Share, 2004). We found that our measure of oral language related to a combined measure of reading comprehension and reading fluency. However, our study did not include direct measures of code-related skills, and determining whether statistical learning might differentially predict precursors of word recognition relating to letter naming, phoneme awareness, or grapheme-to-phoneme correspondences, and if so whether statistical learning ability also predicts reading via these code-related skills, would be a useful future direction for this research. Given that code-related skills relate in part to visual recognition of letters and learning of associations between letters and sounds, there may be a stronger possibility for visual statistical learning ability, in addition to the oral statistical learning tasks that we employed, to relate to reading via this aspect of the model (Arciuli, 2018; Chetail, 2017; Siegelman, Rueckl, Lo, et al., 2022).

Conclusion

We have shown that the well-established link between children's oral language skills in preschool and reading skills in formal schooling is, to a certain extent, underwritten by one aspect of statistical learning, namely children's early ability to detect and use abstract relations among stimuli in speech and abstract over those relations. However, the effect on reading was expressed via oral language, with no evidence of a direct relationship between early statistical learning ability and later reading.

Acknowledgments

This work was supported by the International Centre for Language and Communicative Development (LuCiD) at Lancaster University and the University of Liverpool, funded by the Economic and Social Research Council (UK) (ES/S007113/1).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jecp.2024.106002.

References

Arciuli, J. (2018). Reading as statistical learning. *Language, Speech, and Hearing Services in Schools, 49*(3, Suppl), 634–643.
Arciuli, J., & Simpson, I. C. (2012). Statistical learning is related to reading ability in children and adults. *Cognitive Science, 36*(2), 286–304.
Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238–246.
Bishop, D. (2003). *Test for Reception of Grammar, Version 2 (TROG–2).* Pearson Assessment.
Black, A. W., Taylor, P., & Caley, R. (1990). *The festival speech synthesis system.* Centre for Speech Technology Research, University of Edinburgh. Retrieved from http://www.cstr.ed.ac.uk/projects/festival.html.
Bogaerts, L., Siegelman, N., & Frost, R. (2021). Statistical learning and language impairments: Toward more precise theoretical accounts. *Perspectives on Psychological Science, 16*(2), 319–337.
Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and me: Familiar names help launch babies into speech-stream segmentation. *Psychological Science, 16*(4), 298–304.
Catts, H. W., Adlof, S. M., & Weismer, S. E. (2006). Language deficits in poor comprehenders: A case for the simple view of reading. *Journal of Speech, Language, and Hearing Research, 49*, 278–293.
Chang, Y. N., & Monaghan, P. (2019). Quantity and diversity of preliteracy language exposure both affect literacy development: Evidence from a computational model of reading. *Scientific Studies of Reading, 23*(3), 235–253.
Chang, Y. N., Monaghan, P., & Welbourne, S. (2019). A computational model of reading across development: Effects of literacy onset on language processing. *Journal of Memory and Language, 108*, 104025.
Chetail, F. (2017). What do we do with what we learn? Statistical learning of orthographic regularities impacts written word processing. *Cognition, 163*, 103–120.
Christiansen, M. H. (2019). Implicit-statistical learning: A tale of two literatures. *Topics in Cognitive Science, 11*, 468–481.
Cooper, S. R., Gonthier, C., Barch, D. M., & Braver, T. S. (2017). The role of psychometrics in individual differences research in cognition: A case study of the AX-CPT. *Frontiers in Psychology, 8*, 1482.
Cunillera, T., Cimara, E., Laine, M., & Rodiguez-Fornells, A. (2010). Words as anchors: Known words facilitate statistical learning. *Experimental Psychology, 57*(2), 134–141.
Duff, F. J., Reen, G., Plunkett, K., & Nation, K. (2015). Do infant vocabulary skills predict school-age language and literacy outcomes? *Journal of Child Psychology and Psychiatry, 56*, 848–856.
Dunn, L. M., Dunn, D. M., Sewel, J., & Styles, B. (2009). *The British Picture Vocabulary Scale (BPVS-3)* ((3rd ed.)). GL Assessment.
Evans, J., Saffran, J., & Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 52*, 321–335.
Forest, T. A., Schlichting, M. L., Duncan, K. D., & Finn, A. S. (2023). Changes in statistical learning across development. *Nature Reviews Psychology, 2*(4), 205–219.
Frost, R. (2012). Towards a universal model of reading. *Behavioral and Brain Sciences, 35*(5), 263–279.
Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences, 19*(3), 117–125.
Frost, R. L. A., Jessop, A., Durrant, S., Peter, M. S., Bidgood, A., Pine, J. M., Rowland, C. F., & Monaghan, P. (2020). Non-adjacent dependency learning in infancy, and its link to language development. *Cognitive Psychology, 120*, 101291.
Frost, R. L. A., & Monaghan, P. (2016). Simultaneous segmentation and generalisation of non-adjacent dependencies from continuous speech. *Cognition, 147*, 70–74.
Frost, R. L. A., Monaghan, P., & Christiansen, M. H. (2019). Mark my words: High frequency marker words impact early stages of language learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*(10), 1883–1898.
Frost, R., Siegelman, N., Narkiss, A., & Afek, L. (2013). What predicts successful literacy acquisition in a second language? *Psychological Science, 24*(7), 1243–1252.
Gerbrand, A., Gredebäck, G., Hedenius, M., Forsman, L., & Lindskog, M. (2022). Statistical learning in infancy predicts vocabulary size in toddlerhood. *Infancy, 27*(4), 700–719.
Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. C. (2005). Hard words. *Language Learning and Development, 1* (1), 23–64.
Gómez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition, 70*, 109–135.
Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*(1), 6–10.
Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review, 111*, 662–720.
Hjetland, H. N., Brinchmann, E. I., Scherer, R., Hulme, C., & Melby-Lervåg, M. (2020). Preschool pathways to reading comprehension: A systematic meta-analytic review. *Educational Research Review, 30*, 100323.
Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing, 2*, 127–160.
Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55.
Isbilen, E. S., & Christiansen, M. H. (2022). Statistical learning of language: A meta-analysis into 25 years of research. *Cognitive Science, 46*(9)e13198.
Kemény, F., & Lukács, Á. (2021). The role of statistical learning and verbal short-term memory in impaired and typical lexical development. *Frontiers in Communication, 6*, 700452.
Kendeou, P., van den Broek, P., White, M., & Lynch, J. S. (2009). Predicting reading comprehension in early elementary school: The independent contributions of oral language and decoding skills. *Journal of Educational Psychology, 101*, 765–778.
Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology, 48*, 171–184.
Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual differences in language acquisition and processing. *Trends in Cognitive Sciences, 22*(2), 154–169.
Kim, Y. S., Otaiba, S. A., Puranik, C., Folsom, J. S., & Gruelich, L. (2014). The contributions of vocabulary and letter writing automaticity to word reading and spelling for kindergartners. *Reading and Writing, 27*, 237–253.

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.
Language and Reading Research Consortium & Chiu, Y-D. (2018). The simple view of reading across development: The prediction of Grade 3 reading comprehension by prekindergarten skills. *Remedial and Special Education, 39,* 289–303.
Lee, J. (2011). Size matters: Early vocabulary as a predictor of language and literacy competence. *Applied Psycholinguistics, 32*(1), 69–92.
Lelonkiewicz, J. R., Ktori, M., & Crepaldi, D. (2020). Morphemes as letter chunks: Discovering affixes through visual regularities. *Journal of Memory and Language, 115*, 104152.
Lelonkiewicz, J. R., Ktori, M., & Crepaldi, D. (2023). Morphemes as letter chunks: Linguistic information enhances the learning of visual regularities. *Journal of Memory and Language, 130*, 104411.
Marchetto, E., & Bonatti, L. L. (2015). Finding words and word structure in artificial speech: The development of infants' sensitivity to morphosyntactic regularities. *Journal of Child Language, 42*(4), 873–902.
Metsala, J. L., & Walley, A. C. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 89–120). Lawrence Erlbaum.
Mintz, T. H. (2002). Category induction from distributional cues in an artificial language. *Memory & Cognition, 30*(5), 678–686.
Monaghan, P., Donnelly, S., Alcock, K., Bidgood, A., Cain, K., Durrant, S., Frost, R. L. A., Jago, L. S., Peter, M. S., Pine, J. M., Turnbull, H., & Rowland, C. F. (2023). Learning to generalise but not segment an artificial language at 17 months predicts children's language skills 3 years later. *Cognitive Psychology, 147*, 101607.
Muter, V., Hulme, C., Snowling, M., & Stevenson, J. (2004). Phonemes, rimes, vocabulary and grammatical skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology, 40*, 665–681.
Nation, K., & Snowling, M. J. (2004). Beyond phonological skills: Broader language skills contribute to the development of reading. *Journal of Research in Reading, 27*(4), 342–356.
Oliveira, C. M., Henderson, L. M., & Hayiou-Thomas, M. E. (2023). Limited evidence of an association between language, literacy, and procedural learning in typical and atypical development: A meta-analysis. *Cognitive Science, 47*(7)e13310.
Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology, 98*, 554–566.
Ouellette, G., & Beers, A. (2010). A not-so-simple view of reading: How oral vocabulary and visual-word recognition complicate the story. *Reading and Writing, 23*, 189–208.
Ouellette, G., & Fraser, J. R. (2009). What exactly is a yait anyway: The role of semantics in orthographic learning. *Journal of Experimental Child Psychology, 104*, 239–251.
Panter, A. T., Tanaka, J. S., & Wellens, T. R. (1992). The psychometrics of order effects. In N. Schwarz & S. Sudman (Eds.), *Context effects in social and psychological research* (pp. 249–264). Springer.
Peña, M., Bonatti, L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science, 298*, 604–607.
Perruchet, P., Tyler, M. D., Galland, N., & Peereman, R. (2004). Learning non-adjacent dependencies: No need for algebraic-like computations. *Journal of Experimental Psychology: General, 133*(4), 573–583.
Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language, 39*, 246–263.
Peter, M., Durrant, S., Jessop, A., Bidgood, A., Pine, J., & Rowland, C. F. (2019). Does speed of processing or vocabulary size predict later language growth in toddlers? *Cognitive Psychology, 115*, 101238.
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.
Raviv, L., & Arnon, I. (2018). The developmental trajectory of children's auditory and visual statistical learning abilities: Modality-based differences in the effect of age. *Developmental Science, 21*(4)e12593.
Ren, J., & Wang, M. (2023). Development of statistical learning ability across modalities, domains, and languages. *Journal of Experimental Child Psychology, 226*, 105570.
Ren, J., Wang, M., & Arciuli, J. (2023). A meta-analysis on the correlations between statistical learning, language, and reading outcomes. *Developmental Psychology, 59*(9), 1626–1644.
Renfrew, C. E. (2010). *Bus Story Test: A test of narrative speech*. Speechmark.
Renfrew, C. E., & Mitchell, P. (2010). Word Finding Vocabulary Test. *Speechmark*.
Ricketts, J., Davies, R., Masterson, J., Stuart, M., & Duff, F. J. (2016). Evidence for semantic involvement in regular and exception word reading in emergent readers of English. *Journal of Experimental Child Psychology, 150*, 330–345.
Roembke, T. C., Simonetti, M. E., Koch, I., & Philipp, A. M. (2023). What have we learned from 15 years of research on cross-situational word learning? A focused review. *Frontiers in Psychology, 14*, 1175272.
Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36.
Roth, F. P., Speece, D. L., & Cooper, D. H. (2002). A longitudinal analysis of the connection between oral language and early reading. *Journal of Educational Research, 95*(5), 259–272.
Rowland, C. F., Bidgood, A., Durrant, S., Peter, M., & Pine, J. M. (2024). *The Language 0–5 Project.* Unpublished manuscript, University of Liverpool. Available from https://osf.io/kau5f.
Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science, 12*, 110–114.
Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926–1928.
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.
Share, D. L. (2004). Orthographic learning at a glance: On the time course and developmental onset of self-teaching. *Journal of Experimental Child Psychology, 87*, 267–298.
Siegelman, N., Rueckl, J. G., Lo, J. C. M., Kearns, D. M., Morris, R. D., & Compton, D. L. (2022). Quantifying the regularities between orthography and semantics and their impact on group- and individual-level behavior. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 48*(6), 839–855.

Siegelman, N., Rueckl, J. G., van den Bunt, M., Frijters, J. C., Zevin, J. D., Lovett, M. W., & Morris, R. D. (2022). How you read affects what you gain: Individual differences in the functional organization of the reading system predict intervention gains in children with reading disabilities. *Journal of Educational Psychology, 114*(4), 855–869.

Silva, M. T., & Cain, K. (2015). The relations between lower- and higher-level oral language skills and their role in prediction of early reading comprehension. *Journal of Educational Psychology, 107*, 321–331.

Smith, A. C., Monaghan, P., & Huettig, F. (2021). The effect of orthographic systems on the developing reading system: Typological and computational analyses. *Psychological Review, 128*(1), 125–159.

Snowling, M. J., Stothard, S. E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., & Hulme, C. (2009). *York Assessment of Reading for Comprehension*. GL Assessment.

St. Clair, M. C., Monaghan, P., & Christiansen, M. H. (2010). Learning grammatical categories from distributional cues: Flexible frames for language acquisition. *Cognition, 116*(3), 341–360.

Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology, 38*(6), 934–947.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2012). *TOWRE-2 examiner's manual.* Pro-Ed.

van Witteloostuijn, M., Boersma, P., Wijnen, F., & Rispens, J. (2021). The contribution of individual differences in statistical learning to reading and spelling performance in children with and without dyslexia. *Dyslexia, 27*(2), 168–186.

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.

Wang, F. H., Zevin, J., & Mintz, T. H. (2019). Successfully learning non-adjacent dependencies in a continuous artificial language stream. *Cognitive Psychology, 113*, 101223.

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.

Wilson, B., Spierings, M., Ravignani, A., Mueller, J. L., Mintz, T. H., Wijnen, F., van der Kant, A., Smith, K., & Rey, A. (2020). Non-adjacent dependency learning in humans and other animals. *Topics in Cognitive Science, 12*(3), 843–858.