Individual differences in statistical learning predict children’s reading ability in a semi-transparent orthography

Janne von Koss Torkildsen¹, Joanne Arciuli², and Ona Bø Wie¹,³

¹ University of Oslo, P.O. Box 1140 Blindern, 0318 Oslo, Norway;
² University of Sydney, Cumberland Campus, Lidcombe, NSW 2141, Australia;
³ Oslo University Hospital, Postboks 4950 Nydalen, 0424 Oslo, Norway

Correspondence concerning this article should be addressed to:
Janne von Koss Torkildsen,
Faculty of Educational Sciences,
University of Oslo, Postboks 1140 Blindern, 0318 Oslo, Norway
E-mail: janneto@isp.uio.no

Acknowledgments

This work was supported by the Norwegian Directorate of Health, the University of Oslo and Oslo University Hospital. Joanne Arciuli was supported by a mid-career research fellowship from the Australian Research Council. We wish to thank all the participants in the study and the people who helped collecting the data: Marit Gismarvik, Åsrun Valberg, Ellen Brinchmann, Hanne Røe-Indregård, and MA students at the Department of Special Needs Education, University of Oslo.
Abstract

Learning to read is a milestone in a child’s life, and reading ability is a strong predictor of academic outcomes. Some studies have revealed that individual differences in the capacity for implicit statistical learning are linked with children’s reading skills in English, which has a deep orthography, but we do not know whether the same relation is found in semi-transparent orthographies such as Norwegian. Additionally, the relative contribution of statistical learning alongside other cognitive measures known to be related to reading has not been examined. This study of sixty-five Norwegian school children (\(M = 10.3\) years) found that performance on an independent statistical learning task predicted a small, but unique amount of variance in reading ability in Norwegian, even when a number of other reading-related cognitive skills were taken into account. This suggests that children’s implicit learning of statistical regularities is associated with proficiency with written language in a semi-transparent orthography such as Norwegian.

*Keywords*: learning; literacy; reading acquisition; implicit learning; statistical learning
1. Introduction

Statistical learning (SL), our ability to detect implicitly regularities in the environment, is present from infancy (Kirkham, Slemmer, & Johnson, 2002; Saffran, Aslin, & Newport, 1996), and can be observed in neonates as soon as 2-5 days after birth (Bulf, Johnson, & Valenza, 2011; Teinonen, Fellman, Näätänen, Alku, & Huotilainen, 2009). Further, studies with both children and adults have demonstrated that SL can operate without instruction to learn or remember and without any kind of reinforcement or feedback (Arciuli, Torkildsen, Stevens, & Simpson, 2014; Bertels, Destrebecqz, & Franco, 2015; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997). There is great interest in how individual differences in this kind of implicit learning relate to variability in language proficiency during child development (e.g. Arciuli, 2017; Erickson & Thiessen, 2015; Kidd, Donnelly, & Christiansen, 2018). Recently there has also been growing interest in bringing together the literatures on SL and reading acquisition (e.g. reviews by Arciuli, 2018; Sawi & Rueckl, 2018). This interest is motivated by the idea that learning to read relies, in part, on the ability to detect implicitly statistical regularities in mappings between phonological, lexical-semantic and orthographic forms, and also regularities within each of these domains, such as common clusters of letters (Harm & Seidenberg, 1999; Rueckl, 2016; Seidenberg & McClelland, 1989).

As explained by Arciuli (2018), it is impossible of convey explicitly all of the regularities that children need to acquire in a deep orthography, and this can be demonstrated by considering lesser-known probabilistic relationships between orthography and sound such as orthographic cues to lexical stress. Recent studies indicate that there are numerous probabilistic mappings between letters and sounds, and that these may be learned implicitly, without any instruction. For example, research with English-speaking adults by Arciuli and
CHILDREN’S STATISTICAL LEARNING AND READING

Cuppes (2006, 2007) demonstrated via corpus analyses and behavioral testing that particular groupings of letters within polysyllabic words are probabilistically associated with particular stress patterns in a way that enables the reader to emphasize the correct syllable when reading aloud. Arciuli & Cuppes (2006) found that 91% of disyllabic words ending in “–an” and 64% of words ending in “–ure” exhibit a pattern of first-syllable stress. However, 72% of disyllabic words ending in “–ose” and 86% of words ending in “–uct” exhibit a pattern of second-syllable stress. For a cross-linguistic comparison examining probabilistic orthographic cues to lexical stress in six languages see Monaghan, Arciuli and Seva (2016). A study of English-speaking children by Arciuli, Monaghan and Seva (2010) demonstrated that these probabilistic orthographic cues appear in age-appropriate written texts and are gradually learned by children in a way that enables them to assign lexical stress in predictable ways when reading aloud polysyllabic words and nonwords. Developmental computational modeling confirmed that these statistical patterns can be learned from mere exposure to input containing the regularities.

In line with these results, Apfelbaum, Hazeltine and McMurray (2013) showed that children can learn grapheme-phoneme correspondences (GPCs) without any explicit instruction. The children learned through exposure in computer tasks that included words that embodied the regularities. While it is clearly of benefit to provide beginning readers with tools to assist in ‘cracking the orthographic code’ via systematic explicit instruction in some common GPCs, it is impossible to teach all GPCs explicitly. Thus, it may be assumed that many of these correspondences are learned implicitly. There is also evidence from non-alphabetic languages showing that SL supports children’s reading acquisition. Yin and McBride (2015) found that Chinese children showed sensitivity to regularities in Chinese characters at age 4, before exposure to formal reading instruction, and that these sensitivities predicted word reading a year later. He and Tong (2017) extended this work by showing that
The above studies have examined the statistical properties of mappings between letters and sounds and how sensitivity to these regularities these influence reading acquisition. Another question which has been raised by a few recent studies is whether there is a relation between individual differences in the capacity to detect statistical regularities more generally and word reading abilities (Arciuli & Simpson, 2012; Frost, Siegelman, Narkiss, & Afek, 2013; Nigro, Jiménez-Fernández, Simpson, & Defior, 2015). These studies have examined the relation between reading ability and an independent test of SL which does not use letters as stimuli, but other entities such as sequences of shapes or images. Thus, an underlying assumption in these studies is that although performance on different tasks that involve SL are modality and stimulus dependent, there is also a common set of computational mechanisms devoted to learning of distributional information across different domains and stimulus types (Frost, Armstrong, Siegelman, & Christiansen, 2015). In line with the assumption of a domain-general SL mechanism, a number of studies have demonstrated relations between an independent test of SL and oral language abilities (Kidd, 2012; Kidd & Arciuli, 2016; Misyak & Christiansen, 2012). There also appears to be high correlations between performance on SL tasks in the visual and auditory modalities when participants do not come to the task with prior knowledge regarding the co-occurrence of elements, indicating that there are common computational processes across the two modalities (Siegelman, Bogaerts, Elazar, Arciuli, & Frost, 2018).

In the handful of studies to date that have examined how individual differences on an independent test of SL relate to individual differences in reading skills, the results are somewhat mixed (Arciuli & Simpson, 2012; Frost et al., 2013; Nigro et al., 2015; Qi, Sanchez...
CHILDREN’S STATISTICAL LEARNING AND READING

Araujo, Georgan, Gabrieli, & Arciuli, 2018). In the first of these studies, Arciuli and Simpson (2012) found a relationship between performance on a visual SL task and word reading accuracy in English in typically developing children as well as adults. Frost et al. (2013) examined the role of SL in adult English-speakers’ learning of Hebrew as a second language. Results showed that performance on a visual SL task correlated with degree of improvement in reading skills and learning of word structure. The recent study by Qi et al. (2018) found that auditory SL was significantly related to sentence reading fluency in a sample of English-speaking children and adults. While the results of these studies all indicate that a stimulus-independent SL mechanism plays a role in reading acquisition, the results of a study of Spanish third-graders pointed to a different conclusion. Nigro, Jiménez-Fernández, Simpson, and Defior (2015) did not find a significant association between performance on a visual SL task and word reading accuracy. The authors suggested that this null finding could be due to the fact that Spanish has a transparent orthography. As the one-to-one mappings between orthography and phonology in Spanish are easily taught in schools via explicit instruction, this may reduce the influence of individual differences in implicit learning during children’s reading acquisition. However, the authors also acknowledged that with only 26 participants the study had low statistical power and could not reliably detect moderate effect sizes.

Further evidence regarding a possible link between independent tests of SL and reading skills comes from studies of implicit learning which compare the performance of individuals with and without dyslexia. These studies have used different tasks, such as the serial reaction time (SRT) task and the artificial grammar learning (AGL) task. SRT tasks measure visuomotor sequence learning, while AGL tasks measure learning of a set of rules for how (auditory or visual) elements in a sequence can combine. Common to all these tasks, as well as the more classic SL tasks (e.g. Fiser & Aslin, 2002; Saffran et al., 1996), is that they involve implicit learning of sequential regularities and are considered a type of procedural
CHILDREN'S STATISTICAL LEARNING AND READING

learning (for descriptions of different procedural learning tasks, see Krishnan, Watkins, & Bishop, 2016). A meta-analysis of SRT studies concluded that children and adults with dyslexia show impaired implicit learning compared to controls (Lum, Ullman, & Conti-Ramsden, 2013). The authors suggested that age might play a role in the effect sizes obtained across studies in that differences between dyslexic and control groups seemed to be larger in child studies rather than studies with adults. Another meta-analysis, focusing on implicit visual AGL came to the same conclusion, although questions about publication bias were also raised (van Witteloostuijn, Boersma, Wijnen, & Rispens, 2017). One limitation of the dyslexia research has been the focus on group comparisons. A recent study which took an individual differences approach, found that individuals with dyslexia performed less accurately than controls on tasks measuring auditory SL, and also discovered a positive correlation between the capacity for SL and word reading ability (Gabay, Thiessen, & Holt, 2015).

As there are a number of predictors of children’s reading ability, it is noteworthy that no previous research examining the predictive power of an independent test of SL in explaining variability in children’s reading has taken into consideration individual differences in other cognitive and linguistic skills (although see Qi et al. 2018). This is a necessary next step to gain information about whether SL has predictive power that goes above and beyond the cognitive factors that are typically measured in studies of reading. The present study aimed to fill this gap in the literature by assessing whether a commonly used test of SL would explain unique variance in children’s word reading ability in a design including standardized comprehensive measures of expressive and receptive general language skills, vocabulary, rapid automatized naming (RAN), verbal working memory, non-verbal IQ, and sustained attention. Measures of oral language skills were included as a large number of studies have shown associations between reading abilities and general oral language skills (e.g. Nation &
CHILDREN’S STATISTICAL LEARNING AND READING

Snowling, 2004; Storch & Whitehurst, 2002). Previous studies of elementary school children have found that oral vocabulary contributes not only to reading comprehension, but also to word reading efficiency, which is the reading variable examined in the current study (Cain, 2015; Ouellette & Beers, 2010). Another important skill to assess in the test battery is RAN, due to the large body of studies which have shown a robust link between RAN and reading abilities in a variety of languages, though the specific mechanisms through which it influences reading are still debated (Landerl et al., in press; Lervåg & Hulme, 2009; Norton & Wolf, 2012). Further, verbal working memory has been claimed to play a crucial role in reading development (e.g. Arrington, Kulesz, Francis, Fletcher, & Barnes, 2014; Dixon, LeFevre, & Twilley, 1988), though some studies have found only weak correlations between verbal working memory and reading ability (e.g. Dufva, Niemi, & Voeten, 2001). In recent years, there has been increasing interest in the relationship between visual attention skills and reading, and claims have been made that individual differences in selective and sustained attention influence reading fluency and can predict future reading skills (e.g. Franceschini, Gori, Ruffino, Pedrolli, & Facoetti, 2012; Mishra, 2015). Based on this diverse literature on the cognitive skills that are associated with reading, we included standardized measures of all of the above abilities in the current study.

A critical question relating to the link between SL and language development, be it focused on children’s spoken or written language, is whether the link can be observed across languages that differ in important ways. Cross-linguistic research allows us to investigate particular hypotheses about the relationship between SL and language development and should not be seen as mere ‘replication’. In the current study, we wanted to know whether the link between SL and children’s reading would emerge in Norwegian, which is a semi-transparent orthography. Some languages, such as Finnish and Spanish, are described as having transparent orthographies in the sense that there is a one-to-one relationship between
graphemes and phonemes (e.g., Elley, 1992). Other languages, such as English and unvowelized Hebrew, are considered to have deep orthographies as they have more complex mappings between graphemes and phonemes. In between these extremes are languages with semi-transparent orthographies such as Norwegian, German, and Greek. The GPCs in more transparent orthographies are fewer in number and more easily taught via formal instruction; thus, children’s reading development in these languages may not rely as heavily on implicit mechanisms such as SL. At the same time it is thought that SL operates on any stimuli containing regularities in a way that has been described as ‘automatic’, ‘incidental’ and ‘spontaneous’ (Arciuli & Torkildsen, 2012). Consistent with this, studies have shown that connectionist models relying on statistical learning principles can be applied to languages that differ widely in orthographic depth and other surface properties, suggesting that similar types of probabilistic learning contribute to reading cross-linguistically (Ikeda, Ueno, Ito, Kitagami, & Kawaguchi, 2017; Yang, McCandliss, Shu, & Zevin, 2009). Thus, it could be argued that a higher capacity for SL is likely to boost a child’s reading acquisition in any language, regardless of the number and nature of the GPCs and subsequent ease of explicit instruction.

Norwegian, the language examined in the present study, has approximately 40 phonemes, and an alphabet comprised of 29 letters. While the majority of phonemes are represented by one, and only one, letter, there are a number of non-transparent relations between sounds and letters. These non-transparent relations are not completely random. Rather, there are some regularities in how non-transparent mappings are applied. For example, the sound /œ/ is normally spelled with the letter combination ‘kj’ (as in ‘kjøkken’ [kitchen] and ‘kjole’ [dress]), but only the letter ‘k’ is used to represent this sound when the subsequent vowel is ‘i’ or ‘y’ (as in ‘kino’ [cinema] or ‘kylling’ [chicken]). Thus, a word starting with the letter ‘k’ should be read with a /k/ or a /œ/ sound, depending on the subsequent vowel. Another example is the letter ‘o’ which is read as either /u/ or /ɔ/ depending on whether it
represents a short or a long vowel, and thus depends on mappings between oral lexical knowledge and letters. It is not known whether individual differences in the capacity to detect implicitly statistical regularities are related to children’s reading proficiency in Norwegian.

In summary, the present study examined whether an independent test of children’s capacity for SL predicts their word reading skills in Norwegian. One sub-aim of this study was to investigate whether SL would explain a unique amount of variance in children’s reading ability when individual differences in other critical developmental cognitive and linguistic skills were taken into account. To our knowledge, no previously published study has examined the predictive power of SL on reading when taking into account other cognitive skills commonly associated with reading. Another sub-aim of our study was to explore the link between SL and reading in children who are learning a semi-transparent orthography. We hypothesized that if SL is related to reading ability in a semi-transparent orthography such as Norwegian, it would emerge a significant predictor even when other linguistic and cognitive skills are considered.

2. Methods

2.1. Participants

The initial study of the relation between performance on a test of SL and reading in children found a medium-sized effect of .33 (Arciuli & Simpson, 2012). Based on this previous research, the sample size of the present study was determined by a power analysis, showing that 52 participants would provide adequate statistical power (.80) to detect a medium-sized effect of .33 (G*Power software; Faul, Erdfelder, Buchner, & Lang, 2009). Sixty-seven participants were recruited, but data from two participants was excluded from analyses as these children struggled to understand task instructions. Inclusion criteria were: age between 7 and 12; 11, attending second to seventh grade, Norwegian as native language,
CHILDREN’S STATISTICAL LEARNING AND READING

no known cognitive impairments or diagnoses affecting language or literacy, normal or corrected-to-normal vision according to parental report. All parents of participating children were required to complete a questionnaire including questions about the child’s and parents’ native languages and difficulties relating to language, literacy and cognition. To participate in the study, our inclusion criteria stipulated that the child and at least one of the parents had Norwegian as their native language, and that Norwegian was the main language spoken at home. Further, children with known intellectual disability, dyslexia or language disorder were excluded from the study. Results on tests of nonverbal intelligence, language and reading (see description of tests below) were consistent with results from the parental questionnaire, indicating that there were no participants with intellectual disability, dyslexia or language disorders in the sample. The age range was chosen to be comparable to the study by Arciuli and Simpson (2012) which used the same SL task to assess the relation between word reading and SL. However, where the 38 children in the Arciuli and Simpson study ranged in age from 6;4 to 12;5, we set the lower age limit in the present study to 7 and attending the second semester of second grade, as children in Norway begin reading instruction later than Australian children, and we wanted to ensure that all children had received at least three semesters of reading instruction. The final sample consisted of sixty-five children (41 girls) attending second to seventh grade, ranging in age from 7;2 to 12;11 years ($M_{age} = 10.3, SD_{age} = 1.8$), recruited from public schools in Eastern Norway. An overview of the number of participants per grade level is shown in Table 1. Eight of the 65 children had one parent with another native language than Norwegian.

Parents of participating children had an education level which was higher than the population average in Norway. According to Statistics Norway (2015), 57% of Norwegian women and 40% of Norwegian men in the relevant age group have a higher (post-secondary)
CHILDREN'S STATISTICAL LEARNING AND READING

education. For parents in the present sample, 80% of mothers and 88% of fathers had a post-secondary education.

2.2. Materials and procedure

The present study was part of a larger project where children completed a battery of neuropsychological tests. Approval to conduct the study was granted from the Norwegian Regional Committees for Medical and Health Research Ethics (REC South East Norway). Informed written consent was obtained from the children’s parents and informed oral consent was obtained from the children. The participants were tested in a quiet room at their school or at the Oslo University Hospital. The test battery was administered by the Norwegian-speaking authors [names removed for review purposes] and research assistants, PhD and MA students in speech-language pathology and language and hearing sciences who had received extensive training on the administration of the tests.

2.2.1. Cognitive and Linguistic Predictors of Reading Ability

Nonverbal IQ. To measure general nonverbal IQ, we used Raven's Colored Progressive Matrices (CPM) for children aged 7;0-8;11 and Raven’s Standard Progressive Matrices (standard version or plus version) for children aged 9;0-12;11 (Raven & Court, 2003).

Visual attention. The subtest Attention Sustained (AS) from the Leiter International Performance Scale- Revised (Roid, Miller, & Lieter, 1997) was used to measure visual attention. The child was presented with a collection of geometric shapes on a page, with one or target shape or shape combination at the top of the page. The child was then instructed to cross out as many of the target shapes as possible within the given time. The test consisted of four pages, with increasingly complex shape combinations.
Expressive and Receptive Language. General language skills were assessed with the Norwegian adaptation of the Clinical Evaluation of Language Fundamentals (CELF-4: Semel, Wiig, & Secord, 2003). The CELF-4 includes a Receptive Language index and an Expressive Language index – each derived by summing the scaled scores from a combination of subtests. There are two slightly different versions of the CELF, one for children aged 5-8 and one for children aged 9-12. The Receptive Language Index for both these versions includes a measure of the ability to follow increasingly complex linguistic directions (“After pointing to the shoe, point to the fish.”), and identify semantic similarities between words (indicate which two words in a group of three that fit best with each other). The Expressive Language Index for both versions of the CELF-4 includes a measure of sentence repetition and a measure of the ability to formulate sentences using a specific word (e.g. ‘always’ or ‘neither’) to describe a picture. The subtests RAN and verbal working memory (see below) are not part of the Receptive and Expressive Language Indexes.

Rapid Automatized Naming (RAN). To measure RAN, we used the CELF-4 subscale Rapid Automatic Naming. In this task, the child is presented with a set of visual stimuli containing dual information (4 different color-form combinations) and is asked to rapidly name the pictures seen. The score is the number of seconds used to name all the objects.

Verbal working memory. The digit span subscale on the CELF-4 was used assess verbal working memory. The child is presented with strings of numbers of increasing length and asked to repeat them in the same order as presented or backwards. The score used in the analysis is the composite of forward and backward digit span.

Vocabulary. To assess receptive vocabulary, we used the Norwegian adaptation of the British Picture Vocabulary Scale II (Dunn, Dunn, Whetton, & Berley, 1997; Lyster, Horn, &
CHILDREN'S STATISTICAL LEARNING AND READING

Rygvold, 2010). In each test trial the child is shown four pictures and is asked to indicate which picture that corresponds best to a spoken word.

2.2.2. Visual statistical learning

We used the visual SL task originally reported by Arciuli and Simpson (2011), presented via E-Prime (Psychology Software Tools Inc., 2012). This task utilizes the classic triplet paradigm to assess implicit learning of adjacent dependencies in sequentially presented visual stimuli. The stimuli used in this task are novel cartoon figures (that do not resemble animals, people or fictitious characters), and the task has been used successfully in previous studies examining individual differences in children’s SL (e.g. Arciuli & Simpson, 2012; Kidd & Arciuli, 2016). A recent study by Siegelman et al. (2018) suggested that reliability is higher in SL tasks with non-linguistic stimuli, because these tasks avoid learners’ entrenched expectations based on statistics from their native language experience, and thus they are better suited for studies of individual differences. For examples of previous studies that have used the classic triplet paradigm in the visual modality, but with different stimuli and with adults rather than children, see Brady and Oliva (2008), Fiser and Aslin (2002), and Turk-Browne, Junge, & Scholl (2005).

The SL task is comprised of (1) a familiarization phase (with a cover task) and (2) a surprise test phase. The cartoon figures are described as aliens, and participants are told that the task is a computer game where different types of aliens are queuing to enter a spaceship. As reported in previous studies that have used this task, the aliens are divided into four base triplets, referred to as ABC, DEF, GHI, and JKL (see the appendix in Arciuli & Simpson, 2011). Each triplet occurs 24 times during familiarization (96 presentations of the four triplets). For 6 of the 24 instances, one of the aliens is presented twice in a row. This repetition provides the cover task in that children are required to press the space bar whenever
they detect this repetition. The repetitions are counterbalanced among the three aliens within the triplet (e.g., for base triplet ABC there are occurrences of AABC, ABBC, ABCC, as well as ABC). Familiarization consists of 312 presentations of individual aliens, each appearing 26 times (presentation time of 400 ms, with 200 ms interstimulus interval). The order of the triplets is randomized, following the constraints reported by Turk-Browne et al. (2005) – no repeated triplets (i.e., no instances of ABCABC) and no repeated pairs of triplets (e.g., ABCDEFABCDEF).

In the surprise test phase, learning of triplets is assessed via 64 forced choice trials. Four impossible triplets, each containing one alien from three different base triplets (AEI, DHL, GKC, and JBF), are pitted against base triplets that were embedded in the familiarization stream. On each test trial children are presented with a base triplet and an impossible triplet (order of presentation counterbalanced), using the same presentation rate used during familiarization (with the addition of a 1,000-ms gap separating the base and impossible triplet). After all six aliens have been presented, participants are prompted to identify which of the two triplets had appeared previously in the familiarization phase. There are no time limits regarding responding. A percent correct score is obtained for each child.

2.2.3. Outcome measure: word reading

We used the Norwegian adaptation of the Test of Word Reading Efficiency (TOWRE) (Rashotte, Torgerson, & Wagner, 1999). The Norwegian adaptation of TOWRE has been used in a large number of previous studies of reading development (e.g. Hofslundsengen, Hagtvet, & Gustafsson, 2016; Rydland, Aukrust, & Fulland, 2012; Samuelsson et al., 2008), and there is evidence that it is a reliable and valid test for the age range examined in the present study (Brinchmann, Hjetland, & Lyster, 2016; Lervåg & Aukrust, 2010; Lervåg, Hulme, & Melby-Lervåg, 2018). However, there are no norms available for the Norwegian
adaptation. The TOWRE consists of four word lists, two lists of real words and two lists of nonwords. The words in each list are presented in order of increasing difficulty. The children were instructed to read the lists as fast as they could, and were given 45 seconds to read each list. The number of words or nonwords that were read correctly in the given time was calculated. The variable used in the analysis was the total number of correctly read words from the two lists of real words. Using real word reading as the dependent variable enabled a comparison with the results of Arciuli and Simpson (2012) who also used the same visual SL task in their study of the relation between SL and word reading in English. However, we also ran the regression model with reading performance on all four lists (real and nonword reading) as the outcome variable. The regression model with real and nonword reading as outcome variable was nearly identical to the one reported below, and the percentage of unique variance explained by VSL was in fact slightly higher in this model.

2.2.4 Statistical analyses

For three of the measures used in this study, CELF-4 Expressive Language, CELF-4 Receptive language and Raven’s matrices, children aged 9 years and above received a slightly different version from children who were below 9 years. Thus, we used standard scores from these measures in all analyses. For other measures (i.e. TOWRE, RAN, Leiter AS, BPVS, and SL), children of all ages received the same version, and thus we were able to use raw scores in analyses where age could be taken into account (via partial correlations and regression analyses). To measure the relations between all the variables in the study, we carried out partial correlation analyses, controlling for age. A one sample t-test was used to verify that performance in the SL task was significantly different from chance. Finally, we performed a multiple regression analysis with forced entry to identify the cognitive and linguistic abilities that made a significant contribution to explaining variance in word reading efficiency. For all statistical tests, the alpha level was set to 0.05.
3. Results

3.1. Descriptive statistics and bivariate correlations

Descriptive statistics for the cognitive tests where standard scores are available indicate that this was a typical sample of children with scores close to the norm mean (see Table 2). As can be seen from the partial correlations in Table 3, word reading efficiency was significantly related to performance on the CELF-4 Expressive Language index, verbal working memory, RAN and the SL task. There was also a near significant relation between word reading and visual sustained attention ($r = 0.22$, $p = 0.09$).

Table 1

*Number of participants by grade level*

<table>
<thead>
<tr>
<th>Grade</th>
<th>Number of participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>14</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>20</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
</tr>
</tbody>
</table>

*Note.* All participants were tested in the spring, i.e. their second semester in their grade.
Table 2

*Descriptive Statistics for All Measures*

<table>
<thead>
<tr>
<th>Measure</th>
<th>M</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
<th>Type of score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>10.25</td>
<td>1.76</td>
<td>7.16</td>
<td>12.97</td>
<td>Years</td>
</tr>
<tr>
<td>Word reading</td>
<td>124.08</td>
<td>33.02</td>
<td>43</td>
<td>185</td>
<td>Raw score</td>
</tr>
<tr>
<td>Nonverbal IQ</td>
<td>100.85</td>
<td>13.54</td>
<td>80</td>
<td>140</td>
<td>Standard score</td>
</tr>
<tr>
<td>Visual attention</td>
<td>94.46</td>
<td>12.06</td>
<td>75</td>
<td>130</td>
<td>Standard score</td>
</tr>
<tr>
<td>Receptive language</td>
<td>104.46</td>
<td>11.77</td>
<td>80</td>
<td>134</td>
<td>Standard score</td>
</tr>
<tr>
<td>Expressive language</td>
<td>102.71</td>
<td>10.02</td>
<td>83</td>
<td>125</td>
<td>Standard score</td>
</tr>
<tr>
<td>Verbal WM</td>
<td>91.85</td>
<td>9.62</td>
<td>75</td>
<td>125</td>
<td>Standard score</td>
</tr>
<tr>
<td>RAN</td>
<td>76.95</td>
<td>21.27</td>
<td>46</td>
<td>137</td>
<td>Raw score</td>
</tr>
<tr>
<td>Receptive vocabulary</td>
<td>102.55</td>
<td>14.22</td>
<td>60</td>
<td>126</td>
<td>Standard score</td>
</tr>
<tr>
<td>SL</td>
<td>57.12</td>
<td>13.79</td>
<td>34</td>
<td>97</td>
<td>Per cent correct/Raw score</td>
</tr>
</tbody>
</table>

*Note.* RAN is the number of seconds used to name all the objects in the test, thus a low score signifies better performance. Receptive language and Expressive language are indexes from the CELF-4, and have a mean of 100 and a standard deviation of 15. Verbal WM is the score on the CELF-4 subtest Total digit span, which is comprised of digit span forward and digit span backward. In the CELF-4, Total digit span is a scaled score, but it has been converted to a standard score in this table.
### Table 3

**Partial Correlations between Scores on All Measures (Controlling for Age)**

<table>
<thead>
<tr>
<th>Measure</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Word reading</td>
<td>—</td>
<td>.01</td>
<td>.22</td>
<td>.14</td>
<td>.30*</td>
<td>.44***</td>
<td>-.44***</td>
<td>.05</td>
<td>.30*</td>
</tr>
<tr>
<td>2. Nonverbal IQ</td>
<td>.01</td>
<td>—</td>
<td>-.01</td>
<td>.21</td>
<td>.25*</td>
<td>-.02</td>
<td>-.07</td>
<td>.36**</td>
<td>.28*</td>
</tr>
<tr>
<td>3. Visual attention</td>
<td>.22</td>
<td>-.01</td>
<td>—</td>
<td>.22</td>
<td>.01</td>
<td>.18</td>
<td>-.14</td>
<td>-.11</td>
<td>.05</td>
</tr>
<tr>
<td>5 Expressive lang.</td>
<td>.30*</td>
<td>.25*</td>
<td>.01</td>
<td>.49***</td>
<td>—</td>
<td>.35**</td>
<td>-.34**</td>
<td>.23</td>
<td>.13</td>
</tr>
<tr>
<td>6. Verbal WM</td>
<td>.44***</td>
<td>-.02</td>
<td>.18</td>
<td>.11</td>
<td>.35**</td>
<td>—</td>
<td>-.18</td>
<td>-.07</td>
<td>.10</td>
</tr>
<tr>
<td>7. RAN</td>
<td>-.44***</td>
<td>-.07</td>
<td>-.14</td>
<td>-.19</td>
<td>-.34**</td>
<td>-.18</td>
<td>—</td>
<td>-.20</td>
<td>-.08</td>
</tr>
<tr>
<td>8. Receptive voc.</td>
<td>.05</td>
<td>.36**</td>
<td>-.11</td>
<td>.37**</td>
<td>.23</td>
<td>-.07</td>
<td>-.20</td>
<td>—</td>
<td>-.03</td>
</tr>
<tr>
<td>9. SL</td>
<td>.30*</td>
<td>.28*</td>
<td>.05</td>
<td>.14</td>
<td>.13</td>
<td>.10</td>
<td>-.08</td>
<td>-.03</td>
<td>—</td>
</tr>
</tbody>
</table>

*Note.* *p < 0.05, **p < 0.01, ***p < 0.001. Scores for word reading (TOWRE), visual attention (Leiter R), RAN (CELF-4), receptive vocabulary (BPVS) and SL are raw scores. Scores for Nonverbal IQ (Raven’s Matrices), the Receptive Language Index (CELF-4), the Expressive Language Index (CELF-4) and verbal working memory (CELF-4) are standard scores.
3.2. Statistical learning

The mean proportion of correctly identified triplets in the surprise test phase of the SL task was 57.12 ($SD = 13.79$), which is similar to the performance seen in a previous study using the same task and the same presentation speed with English-speaking children of approximately the same age (56.5%) (Arciuli & Simpson, 2011). A one-sample t-test showed that learning was statistically significantly higher than a chance score of 50 ($t (64) = 4.16$, $p < 0.001$, $d = 0.52$) by 7.12 (95% CI: 3.71 to 10.54). Cronbach’s alpha for the VSL task was 0.81.

3.3. Predictors of word reading

To identify the cognitive and linguistic abilities that made a significant contribution to explaining variance in word reading efficiency, a multiple regression analysis with forced entry was performed, with age, nonverbal IQ, RAN, verbal working memory, visual attention, receptive language, expressive language, vocabulary, and SL as the independent variables (see Table 4). This model explained 64% of the total variance in children’s word reading efficiency. RAN, verbal working memory, and SL were the three variables in the model that made a unique contribution to explaining significant variance in word reading efficiency. Specifically, performing better on the tasks measuring RAN, verbal working memory and SL, were related to more efficient reading of words. The percentage of unique variance in reading efficiency explained by each of significant predictors in the model was 5.3% for RAN, 4.8% for verbal working memory and 4.0% for SL. Note that the remainder is shared variance.

Variance inflation factors for all predictors were below 5.1, which suggests no threat of multicollinearity (Hair, Anderson, Tatham, & Black, 1995). The results of the regression analyses correspond well with the findings in the partial correlation matrix (Table 3). Of the four variables that were significantly correlated with word reading efficiency, SL and the
CHILDREN'S STATISTICAL LEARNING AND READING

CELF-4 Expressive Language Index had the lowest correlations. The Expressive Language Index shared a substantial amount of variance with the two stronger predictors (Verbal WM and RAN), while SL was not significantly correlated with any of these variables. Thus, only Verbal WM, RAN and SL emerged as significant predictors in the regression model.

Table 4

Summary of Multiple Regression Analysis for Variables Predicting Word Reading Efficiency

<table>
<thead>
<tr>
<th>Predictors</th>
<th>B</th>
<th>SE B</th>
<th>B</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age in years</td>
<td>5.52</td>
<td>3.43</td>
<td>.29</td>
<td>1.61</td>
<td>.113</td>
</tr>
<tr>
<td>Nonverbal IQ</td>
<td>-0.24</td>
<td>0.25</td>
<td>-.10</td>
<td>-0.97</td>
<td>.341</td>
</tr>
<tr>
<td>Visual attention</td>
<td>0.19</td>
<td>0.18</td>
<td>.13</td>
<td>1.06</td>
<td>.292</td>
</tr>
<tr>
<td>Receptive lang.</td>
<td>-0.15</td>
<td>0.29</td>
<td>-.05</td>
<td>-0.51</td>
<td>.612</td>
</tr>
<tr>
<td>Expressive lang.</td>
<td>0.26</td>
<td>0.39</td>
<td>.08</td>
<td>0.66</td>
<td>.510</td>
</tr>
<tr>
<td>Verbal WM</td>
<td>4.26</td>
<td>1.55</td>
<td>.25</td>
<td>2.74</td>
<td>.008</td>
</tr>
<tr>
<td>RAN</td>
<td>-0.51</td>
<td>0.18</td>
<td>-.33</td>
<td>-2.81</td>
<td>.007</td>
</tr>
<tr>
<td>Receptive voc.</td>
<td>0.21</td>
<td>0.33</td>
<td>.09</td>
<td>0.65</td>
<td>.521</td>
</tr>
<tr>
<td>SL</td>
<td>0.50</td>
<td>0.21</td>
<td>.21</td>
<td>2.42</td>
<td>.019</td>
</tr>
</tbody>
</table>

Note. B = unstandardized coefficients, SE B = standard error for the unstandardized coefficients, β = standardized coefficients.

4. Discussion

4.1. Statistical learning and reading

In the present study we explored the association between SL and children’s reading ability when a number of reading-related cognitive and linguistic abilities were taken into
CHILDREN'S STATISTICAL LEARNING AND READING

account. We found that children’s capacity for SL predicted variability in their word reading above and beyond other reading-related cognitive skills. Specifically, our results showed that independent tests of SL, RAN and verbal working memory made unique contributions to explaining variance in word reading efficiency. Our final regression model explained 64% of variability in children’s reading ability, which is comparable to other studies of reading in Norwegian (Furnes & Samuelsson, 2010, 2011; Helland & Morken, 2016). It should be noted that the amount of unique variance explained by SL was small, and smaller than the variance explained by RAN and verbal working memory.

Broadly speaking, the SL task used here assesses implicit learning of statistical regularities. As there are statistical regularities present in the way orthography maps to phonology and in the way that letters can be combined within orthography, it follows that that a capacity for SL would be related to the reading aloud of words. However, we cannot draw conclusions about the specific SL mechanisms that account for the relationship between SL and reading in the present study. As mentioned in the introduction, performance in SL tasks may rely on both domain-specific and domain-general abilities (Arciuli, 2017; Frost et al., 2015). Thus, aside from the possibility that our visual SL task measures domain-general computational mechanisms related to reading acquisition, it possible that the task specifically taps visual implicit learning skills which contribute to acquisition of transitional probabilities between letters, such as frequent bigrams and trigrams. Auditory SL tasks, on the other hand, may to a larger extent tap abilities that support learning of phoneme-grapheme correspondences. Thus, employing an auditory task, or a task using linguistic stimuli, may have led to different results. Indeed, a recent study comparing the effects of auditory and visual SL on reading in English, found that auditory SL, but not visual SL, was significantly associated with sentence reading fluency (Qi et al., 2018). Moreover, auditory SL was also significantly related to accuracy of nonword reading in children, a relationship which seemed
CHILDREN’S STATISTICAL LEARNING AND READING

to be mediated by phonological processing skills. However, in addition to the modality difference, there were a number of other differences between the auditory and visual SL tasks used in the study by Qi et al., such as presentation rate and the nature of the cover task, and thus there is a need for further studies that investigate the effects of such methodological factors on the relation between SL and reading. In sum, although there is growing interest in the precise sub-components of SL tasks, these are yet to be determined (for a theoretical framework outlining the multi-component nature of SL, see Arciuli, 2017). However, our results suggest that the SL task used here taps into skills that differ from those assessed by our battery of other cognitive and linguistic tests. Moreover, our analyses show that the SL task used here exhibits high reliability in terms of internal consistency.

It has been suggested that SL may only play a role in learning to read in deep orthographies such as English where the GPCs are complex and difficult to teach via explicit instruction (Nigro et al., 2015). Indeed, previous studies have demonstrated that individual differences in SL are related to reading ability in English and Hebrew (Arciuli & Simpson, 2012; Frost et al., 2013; Qi et al., 2018), but not in Spanish, which has a transparent orthography (Nigro et al., 2015). However, it should be noted that the study of Spanish was underpowered to detect a medium-sized effect, and thus there is a need for further studies investigating the relation between SL and reading skills in transparent orthographies before firm conclusions can be drawn. Results from the present study indicate that SL plays a role in reading acquisition in a semi-transparent orthography such as Norwegian. Moreover, effect sizes for the correlation between SL and children’s reading in Arciuli and Simpson’s study of English and our study of Norwegian were similar (correlation coefficients were .33 and .30 respectively when age was taken into account), suggesting associations of similar magnitude between SL and reading in deep and semi-transparent orthographies. The study by Frost et al. (2013) found somewhat stronger relationships between SL and degree of improvement in
CHILDREN’S STATISTICAL LEARNING AND READING

reading in Hebrew as a second language, which may indicate that SL plays a more important role in second language learning. However, reading *improvement* was the dependent variable in that study, and the participants were adults, which makes it difficult to make direct comparisons.

4.2. Statistical learning and the cognitive and linguistic predictor variables

A number of previous studies have found a link between SL and children’s spoken language development in terms of vocabulary (e.g. Evans, Saffran, & Robe-Torres, 2009) and syntax (e.g. Kidd, 2012; Kidd & Arciuli, 2016). Against this backdrop, it may seem surprising that we did not find a significant relation between SL and expressive or receptive language as measured by the CELF-4 or vocabulary as measured by the BPVS in the present study. However, other studies have failed to observe a link between SL and vocabulary in school-aged children (Kidd & Arciuli, 2016; Kidd & Kirjavainen, 2011; Lum, Conti-Ramsden, Page, & Ullman, 2012; Tomblin, Mainela-Arnold, & Zhang, 2007). Kidd and Arciuli (2016) pointed out that a variety of different learning mechanisms support vocabulary acquisition. They hypothesized that SL may be most strongly linked with vocabulary acquisition during young children’s development, and that this link may weaken as children grow older and other mechanisms begin to influence vocabulary growth. Moreover, with regard to the link between SL and syntax, Kidd and Arciuli (2016) found that SL only predicted variance in performance on particular syntactic structures (passives and object relative clauses but not actives and subject relative clauses). It may be that the standardized tests we used are not sensitive enough to draw out a link with SL in our older school-aged children.

In this study we observed a moderate and statistically significant correlation between performance on the SL task and a test of nonverbal intelligence (Raven’s Matrices). Many previous studies have suggested that SL is separate from intelligence or the *g* factor (Evans et
CHILDREN'S STATISTICAL LEARNING AND READING

al., 2009; Kaufman et al., 2010; Kidd & Arciuli, 2016; Siegelman & Frost, 2015). There are, however, other studies that have observed a significant relation between IQ and performance on implicit learning tasks (Fletcher, Maybery, & Bennett, 2000; Salthouse, McGuthry, & Hambrick, 1999). A possible reason for why this link was observed in the present study could be that both the VSL task and Raven’s matrices rely on maintaining attention to novel visual stimuli (cartoon figures in the VSL task and more abstract figures in Raven’s matrices) over time. However, we included an independent measure of sustained visual attention in the test battery, and neither performance on the nonverbal intelligence task, nor performance on the VSL task was significantly related to this measure. Thus, it may be that the SL task and the test of nonverbal intelligence both rely on a type of attention that we were not able to measure with the test used in this study. Alternatively, there may be other shared underlying factors contributing to performance on both tasks, such as processing speed. A number of studies have suggested a relation between processing speed and intelligence (Kail, 2000), and a relation between processing speed and implicit learning has also been reported (Kaufman et al., 2010). Clarifying the associations between SL and basic cognitive mechanisms such as attention and processing speed may help elucidate the mixed findings regarding the relation between SL and performance on complex cognitive tasks such as intelligence tests.

4.3 Reading and the other cognitive and linguistic predictor variables

Regarding predictors of word reading apart from SL, we found a significant association between reading and measures of expressive language, verbal working memory, and RAN. There was also a near significant relation between word reading and visual sustained attention. These relations were expected based on the previous literature on the association between word reading and expressive language (e.g. Ouellette & Beers, 2010), working memory (e.g. Arrington et al., 2014), RAN (e.g. Landerl et al., in press) and
CHILDREN’S STATISTICAL LEARNING AND READING

sustained attention (e.g. Franceschini et al., 2012). However, there was no correlation between word reading and the receptive language measures from CELF-4 and BPVS. A possible reason for this is that these measures primarily assessed semantic and syntactic skills. While performance on the BPVS and CELF-4 Receptive Language Index depended heavily on knowledge of the meaning of specific words (BPVS) and their semantic and syntactic relation to other words (CELF-4), the reading measure involved lists of single unrelated words and did not assess semantic knowledge. Additionally, the instruction to read the words aloud as fast as possible may have directed attention to phonological rather than semantic aspects of the stimulus words.

4.4. Limitations and future directions

The sample in the present study was relatively small (65 children), but comparable to previous studies of the relation between SL and proficiency in language or reading (e.g. Arciuli & Simpson, 2012; Kidd & Arciuli, 2016; Misyak & Christiansen, 2012). While our study did not include children in the first one and a half years of schooling, a wide age and grade range was also represented. A larger sample and a longitudinal design would have enabled an examination of how the relation between statistical learning and reading may change during the course of reading development. Such a study, which would be an interesting next step, should ideally also use multiple SL tasks in different modalities. For example, it is possible that auditory SL mechanisms are the most critical at the earlier stages of reading development where grapheme-phoneme correspondences are learned, while visual SL mechanisms may support learning of transitional probabilities between letters which may continue to improve even after phoneme-grapheme correspondences are well-established.

Another limitation in the present study is that it did not include a measure of phonological awareness, which is an established predictor of reading skills during the first
CHILDREN'S STATISTICAL LEARNING AND READING stages of reading acquisition and also beyond this stage in deep orthographies (Catts, Gillispie, Leonard, Kail, & Miller, 2002; Torgesen, Wagner, Rashotte, Burgess, & Hecht, 1997). Our decision not to include phonological awareness was based on studies of semi-transparent orthographies which have found either that phonological awareness has only a very slight relationship with reading acquisition (Harris & Giannouli, 1999), or that it is only a reliable predictor in kindergarten and first grade (Aarnoutse, van Leeuwe, & Verhoeven, 2005; Landerl et al., in press; Verhagen, Aarnoutse, & Van Leeuwe, 2008; Wimmer, Mayringer, & Landerl, 2000; Öney & Durgunoğlu, 1997). Recent large-scale studies of Norwegian have shown that RAN makes an important contribution to explaining variance in reading acquisition, but that phonological awareness may not after the first year of reading instruction (Furnes & Samuelsson, 2010, 2011; Helland & Morken, 2016; Lervåg, Bråten, & Hulme, 2009). The present study focused on children from second to seventh grade. Despite this, the inclusion of phonological awareness as a predictor may likely have enabled us to explain more variance in word reading efficiency than our current regression model.

4.5. Conclusions and implications

The finding in the present study that performance on an independent test of implicit SL was associated with word reading efficiency is consistent with the view that reading acquisition happens partly through implicit mechanisms, also in semi-transparent orthographies such as Norwegian. The association between SL and reading was seen even though the SL task did not involve letters or letter-like stimuli, suggesting that stimulus-independent SL skills may be involved. Finally, our results indicate that the SL task taps a reading-related ability which is different from those measured by standard cognitive and linguistic tests which are typically included in studies of the predictors of reading ability.
CHILDREN'S STATISTICAL LEARNING AND READING

Assuming that reading acquisition relies in part on implicit learning, an important question for reading instruction is how we can facilitate this type of learning. Previous research has suggested that the way the input is structured can play a key role for the efficiency of implicit learning mechanisms (for a review, see Plante & Gómez, 2018). Learners implicitly search for regularity in the input. Thus, a way to emphasize specific statistical patterns in language or orthography is to make the target pattern the most consistent event in the learning experience. This can be achieved by introducing high variability in all non-target elements. An example of how this principle can be applied to reading instruction is the study by Apfelbaum, Hazeltine and McMurray (2013) where first-grade students were trained on six regularities between graphemes and phonemes for vowels. Half the students received training input where the vowels were embedded in variable consonant frames, while the other half worked with the same vowels in similar consonant frames. The students did not receive any explicit instruction regarding the regularities. Results showed that variable consonant frames resulted in better learning of the regularities than similar consonant frames. This finding indicated that high variability in non-target elements helped to channel the learners’ attention to the parts of the stimuli that contained relevant statistics. More generally, these findings underline the importance of considering implicit learning mechanisms when developing materials for literacy instruction. As some meta-analyses have indicated that individuals with dyslexia have impairments in implicit learning (Lum et al., 2013; van Witteloostuijn et al., 2017), tailoring the orthographic input in a way that emphasizes statistical regularities may be especially important for this group.

In sum, the present study provides the first evidence suggesting an association between SL and children’s word reading skills in the semi-transparent orthography of Norwegian, even after individual differences in a number of other reading-related cognitive and linguistic skills are taken into account. The link between individual differences in SL and
CHILDREN'S STATISTICAL LEARNING AND READING

reading ability in Norwegian is in line with previous studies of deep orthographies such as English and Hebrew. We hope that our study will encourage more cross-linguistic research on the link between SL and development of both spoken and written language abilities.


CHILDREN'S STATISTICAL LEARNING AND READING


CHILDREN'S STATISTICAL LEARNING AND READING


CHILDREN'S STATISTICAL LEARNING AND READING


